

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 8.11111 Seconds
(without alignments)
27.102 Million cell updates/sec

Title: US-09-905-083-31
Perfect score: 49
Sequence: 1 KMNEYTVHL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	DB ID	Description
1	49	100.0	9	US-09-502-600-31	Sequence 31, Appl
2	49	100.0	9	US-09-502-600-89	Sequence 89, Appl
3	49	100.0	9	US-09-502-600-109	Sequence 109, Appl
4	49	100.0	154	US-09-261-416-7	Sequence 7, Appl
5	49	100.0	224	US-08-944-483-33	Sequence 33, Appl
6	49	100.0	225	US-08-557-146-12	Sequence 12, Appl
7	49	100.0	225	US-09-027-337-4	Sequence 4, Appl
8	49	100.0	225	US-09-154-344-12	Sequence 12, Appl
9	49	100.0	253	US-08-557-146-2	Sequence 2, Appl
10	49	100.0	253	US-08-824-874-3	Sequence 3, Appl
11	49	100.0	253	US-09-154-344-2	Sequence 2, Appl
12	49	100.0	253	US-08-930-188-2	Sequence 2, Appl
13	49	100.0	253	US-09-210-084-3	Sequence 3, Appl
14	49	100.0	253	PCT-US96-04294-2	Sequence 2, Appl
15	39	79.6	9	US-09-502-600-122	Sequence 122, Appl
16	37	75.5	9	US-09-502-600-93	Sequence 93, Appl
17	33	67.3	799	US-08-909-954-4	Sequence 4, Appl
18	33	67.3	804	US-08-909-954-2	Sequence 2, Appl
19	32	65.3	121	US-09-091-725-49	Sequence 49, Appl
20	31	63.3	575	US-08-913-805A-2	Sequence 2, Appl
21	31	63.3	575	US-09-442-629-2	Sequence 2, Appl
22	31	63.3	756	US-08-982-785A-2	Sequence 2, Appl
23	31	63.3	1183	US-08-447-031A-2	Sequence 2, Appl
24	30	61.2	28	PCT-US94-05150-12	Sequence 12, Appl
25	30	61.2	42	PCT-US94-05150-17	Sequence 17, Appl
26	30	61.2	593	US-08-900-927-1	Sequence 1, Appl
27	30	61.2	593	US-08-900-927-3	Sequence 3, Appl

```

28 30 61.2 593 2 US-08-900-927-4 Sequence 4, Appl
29 30 61.2 593 2 US-09-191-279-1 Sequence 1, Appl
30 30 61.2 593 2 US-09-191-279-3 Sequence 3, Appl
31 30 61.2 593 2 US-09-191-279-4 Sequence 4, Appl
32 30 61.2 593 4 US-09-334-476-1 Sequence 1, Appl
33 30 61.2 593 4 US-09-334-476-3 Sequence 3, Appl
34 30 61.2 593 4 US-09-334-476-4 Sequence 4, Appl
35 30 61.2 743 2 US-09-012-030-2 Sequence 2, Appl
36 30 61.2 743 2 US-08-590-454-2 Sequence 2, Appl
37 30 61.2 911 3 US-09-074-579-4 Sequence 4, Appl
38 30 61.2 911 4 US-09-388-774-4 Sequence 4, Appl
39 30 61.2 976 2 US-08-449-645A-18 Sequence 18, Appl
40 30 61.2 976 2 US-08-702-367A-18 Sequence 18, Appl
41 30 61.2 976 5 PCT-US95-04681-18 Sequence 18, Appl
42 30 61.2 977 2 US-08-673-789-8 Sequence 8, Appl
43 30 61.2 2703 1 US-08-185-432-19 Sequence 19, Appl
44 29 59.2 297 1 US-08-534-910B-6 Sequence 6, Appl
45 29 59.2 297 1 US-08-534-910B-7 Sequence 7, Appl

```

ALIGNMENTS

```

RESULT 1
US-09-502-600-31
; Sequence 31, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 31
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 72-80 of the SCCE protein
US-09-502-600-31

```

```

Query Match 100.0%; Score 49; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 1 KMNEYTVHL 9
Db 1 KMNEYTVHL 9

```

RESULT 2

```

US-09-502-600-89
; Sequence 89, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 89
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```

; OTHER INFORMATION: Residues 72-80 of the SCCE protein
US-09-502-600-89

Query Match 100.0%; Score 49; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|||||
Db 1 KMNEYTVHL 9

RESULT 3

US-09-502-600-109
; Sequence 109, Application US/09502600A
; Patent No. 6294344

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; TITLE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 109

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Residues 72-80 of the SCCE protein

US-09-502-600-109

Query Match 100.0%; Score 49; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|||||
Db 1 KMNEYTVHL 9

RESULT 4

US-09-261-416-7
; Sequence 7, Application US/09261416A
; Patent No. 6291663

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14

; SEQ ID NO 7

; LENGTH: 154

; TYPE: PRT

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: Serine protease catalytic domain of stratum corneum

; OTHER INFORMATION: chymotryptic enzyme (Scce) homologous to similar

; OTHER INFORMATION: domain in TADG-12

US-09-261-416-7

Query Match 100.0%; Score 49; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.008;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|||||
Db 18 KMNEYTVHL 26

RESULT 5

US-08-944-483-33
; Sequence 33, Application US/08944483
; Patent No. 6232456

; GENERAL INFORMATION:

; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.

; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS

; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES

; NUMBER OF SEQUENCES: 76

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Abbott Laboratories

; STREET: 100 Abbott Park Road

; CITY: Abbott Park

; STATE: IL

; COUNTRY: USA

; ZIP: 60064-3500

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette

; COMPUTER: IBM Compatible

; OPERATING SYSTEM: DOS

; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/944,483

; FILING DATE:

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER:

; FILING DATE:

; ATTORNEY/AGENT INFORMATION:

; NAME: Becker, Cheryl L.

; REGISTRATION NUMBER: 35,441

; REFERENCE/DOCKET NUMBER: 6183.US.01

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 847/935-1729

; TELEFAX: 847/938-2623

; TELEX:

; INFORMATION FOR SEQ ID NO: 33:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 224 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: No. 6232456e

US-08-944-483-33

Query Match 100.0%; Score 49; DB 4; Length 224;

Best Local Similarity 100.0%; Pred. No. 0.012;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

|||||

Db 43 KMNEYTVHL 51

RESULT 6

US-08-557-146-12

; Sequence 12, Application US/08557146

; Patent No. 5834290

; GENERAL INFORMATION:

; APPLICANT: Egelrud, Torbjorn

; APPLICANT: Hansson, Lennart

; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic

; TITLE OF INVENTION: Enzyme (SCCE)

;; NUMBER OF SEQUENCES: 17
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: White & Case, Patent Department
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036-2787
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/557,146
;; FILING DATE: 14-DEC-1995
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sterner, Richard J.
;; REGISTRATION NUMBER: 35,372
;; REFERENCE/DOCKET NUMBER: 1103326-181
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 819-8783
;; TELEFAX: (212) 354-8113
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 225 amino acids
;; TYPE: amino acids
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: polypeptide
US-08-557-146-12

Query Match 100.0%; Score 49; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|||||
DB 44 KMNEYTVHL 52

RESULT 7
US-09-027-337-4
;; Sequence 4, Application US/09027337B
;; Patent No. 5972616
;; GENERAL INFORMATION:
;; APPLICANT: O'Brien, Timothy J.
;; APPLICANT: Tanimoto, Hiroto
;; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
;; FILE REFERENCE: D6064
;; CURRENT APPLICATION NUMBER: US/09/027,337B
;; CURRENT FILING DATE: 1998-02-20
;; NUMBER OF SEQ ID NOS: 13
;; SEQ ID NO 4
;; LENGTH: 225
;; TYPE: PRT
;; ORGANISM: Unknown
;; FEATURE:
;; OTHER INFORMATION: serine protease catalytic domain of Scce homologous to
;; OTHER INFORMATION: similar domain in TADG-15
US-09-027-337-4

Query Match 100.0%; Score 49; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|||||
DB 44 KMNEYTVHL 52

RESULT 8
US-09-154-344-12
;; Sequence 12, Application US/09154344
;; Patent No. 5981256
;; GENERAL INFORMATION:
;; APPLICANT: Egelrud, Torbjorn
;; APPLICANT: Hansson, Lennart
;; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
;; TITLE OF INVENTION: Enzyme (SCCE)
;; NUMBER OF SEQUENCES: 17
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: White & Case, Patent Department
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036-2787
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/154,344
;; FILING DATE: 16-SEP-1998
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/557,146
;; FILING DATE: 14-DEC-1995
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sterner, Richard J.
;; REGISTRATION NUMBER: 35,372
;; REFERENCE/DOCKET NUMBER: 1103326-181
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 819-8783
;; TELEFAX: (212) 354-8113
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 225 amino acids
;; TYPE: amino acids
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: polypeptide
US-09-154-344-12

Query Match 100.0%; Score 49; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.012;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|||||
DB 44 KMNEYTVHL 52

RESULT 9
US-08-557-146-2
;; Sequence 2, Application US/08557146
;; Patent No. 5834290
;; GENERAL INFORMATION:
;; APPLICANT: Egelrud, Torbjorn
;; APPLICANT: Hansson, Lennart
;; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
;; TITLE OF INVENTION: Enzyme (SCCE)
;; NUMBER OF SEQUENCES: 17
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: White & Case, Patent Department
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: U.S.A.
;; ZIP: 10036-2787
;; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/557,146
;; FILING DATE: 14-DEC-1995
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Sterner, Richard J.
;; REGISTRATION NUMBER: 35,372
;; REFERENCE/DOCKET NUMBER: 1103326-181
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 819-8783
;; TELEFAX: (212) 354-8113
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 253 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-557-146-2

Query Match 100.0%; Score 49; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KMNEYTVHL 9
Db 72 KMNEYTVHL 80

RESULT 10
US-08-824-874-3
; Sequence 3, Application US/08824874
; Patent No. 5962300
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE: Filed Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-835-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

;; IMMEDIATE SOURCE:
;; LIBRARY: GenBank
;; CLONE: 532504
US-08-824-874-3
Query Match 100.0%; Score 49; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KMNEYTVHL 9
Db 72 KMNEYTVHL 80

RESULT 11
US-09-154-344-2
; Sequence 2, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-154-344-2

Query Match 100.0%; Score 49; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KMNEYTVHL 9
Db 72 KMNEYTVHL 80

RESULT 12
US-08-930-188-2
; Sequence 2, Application US/08930188
; Patent No. 6093397
; GENERAL INFORMATION:

```

; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; APPLICATION NUMBER: US/08/930,188
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-930-188-2

Query Match 100.0%; Score 49; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 72 KMNEYTVHL 80

RESULT 13
US-09-210-084-3
; Sequence 3, Application US/09210084
; Patent No. 6197511
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,084

```

```

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 532504
US-09-210-084-3

Query Match 100.0%; Score 49; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 72 KMNEYTVHL 80

RESULT 14
PCT-US96-04294-2
; Sequence 2, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-04294-2

Query Match 100.0%; Score 49; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.014;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KMNEYTVHL 9
|||||
Db 72 KMNEYTVHL 80

RESULT 15

US-09-502-600-122
; Sequence 122, Application US/09502600A
; Patent No. 6294344

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of

; FILE REFERENCE: D6223CIP-C

; CURRENT FILING DATE: 2000-02-11

; CURRENT APPLICATION NUMBER: US/09/502,600A

; PRIOR APPLICATION NUMBER: 09/039,211

; PRIOR FILING DATE: 03-14-1998

; NUMBER OF SEQ ID NOS: 136

; SEQ ID NO 122

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Residues 74-82 of the SCCE protein

US-09-502-600-122

Query Match 79.6%; Score 39; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 NEYTVHL 9
|||||
Db 1 NEYTVHL 7

RESULT 16

US-09-502-600-93
; Sequence 93, Application US/09502600A
; Patent No. 6294344

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of

; FILE REFERENCE: D6223CIP-C

; CURRENT FILING DATE: 2000-02-11

; CURRENT APPLICATION NUMBER: US/09/502,600A

; PRIOR APPLICATION NUMBER: 09/039,211

; PRIOR FILING DATE: 03-14-1998

; NUMBER OF SEQ ID NOS: 136

; SEQ ID NO 93

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Residues 70-78 of the SCCE protein

US-09-502-600-93

Query Match 75.5%; Score 37; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KMNEYTV 7
|||||
Db 3 KMNEYTV 9

RESULT 17

US-08-909-954-4

; Sequence 4, Application US/08909954A

; Patent No. 6100058

; GENERAL INFORMATION:

; APPLICANT: Allen, Maxine J.

; APPLICANT: Buckler, Alan J.

; TITLE OF INVENTION: GAP12 Genes and their Uses

; FILE REFERENCE: SEQ-11P

; CURRENT APPLICATION NUMBER: US/08/909,954A

; CURRENT FILING DATE: 1997-08-12

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 799

; TYPE: PRT

; ORGANISM: M. musculus

US-08-909-954-4

Query Match 67.3%; Score 33; DB 3; Length 799;
Best Local Similarity 100.0%; Pred. No. 87;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 EYTVHL 9
|||||
Db 56 EYTVHL 61

RESULT 18

US-08-909-954-2

; Sequence 2, Application US/08909954A

; Patent No. 6100058

; GENERAL INFORMATION:

; APPLICANT: Allen, Maxine J.

; APPLICANT: Buckler, Alan J.

; TITLE OF INVENTION: GAP12 Genes and their Uses

; FILE REFERENCE: SEQ-11P

; CURRENT APPLICATION NUMBER: US/08/909,954A

; CURRENT FILING DATE: 1997-08-12

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 804

; TYPE: PRT

; ORGANISM: H. sapiens

; FEATURE:

US-08-909-954-2

Query Match 67.3%; Score 33; DB 3; Length 804;
Best Local Similarity 100.0%; Pred. No. 88;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 EYTVHL 9
|||||
Db 56 EYTVHL 61

RESULT 19

US-09-091-725-49

; Sequence 49, Application US/09091725

; Patent No. 6329141

; GENERAL INFORMATION:

; APPLICANT:

; TITLE OF INVENTION: Improved methods for transforming Phaffia

; TITLE OF INVENTION: and recombinant DNA for use therein

; NUMBER OF SEQUENCES: 51

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Morrison & Foerster llp

; STREET: 2000 Pennsylvania Avenue, N.W.

; CITY: Washington

; STATE: DC

COUNTRY: United States of America
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25 (BPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/091,725
FILING DATE: 23-DEC-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95203620.0
FILING DATE: 22-DEC-1995
APPLICATION NUMBER: EP 96200943.7
FILING DATE: 11-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: E. Victor Donahue
REGISTRATION NUMBER: 35,492
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 121 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-091-725-49

Query Match 65.3%; Score 32; DB 4; Length 121;
Best Local Similarity 83.3%; Pred. NO. 17;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9
|||:|
DB 19 EYTIHL 24

RESULT 20
US-08-913-805A-2
Sequence 2, Application US/08913805A
Patent No. 6054304
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naofumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365

FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 575 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-913-805A-2
Query Match 63.3%; Score 31; DB 3; Length 575;
Best Local Similarity 62.5%; Pred. NO. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 MNEYTVHL 9
:|||||
DB 379 IEYTVHV 386
RESULT 21
US-09-442-629-2
Sequence 2, Application US/09442629
Patent No. 6291219
GENERAL INFORMATION:
APPLICANT: TANIGUCHI, Naoyuki
APPLICANT: UOZUMI, Naofumi
APPLICANT: SHIBA, Tetsuo
APPLICANT: YANAGIDANI, Shusaku
TITLE OF INVENTION: Alpha 1-6 Fucosyltransferase
NUMBER OF SEQUENCES: 15
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: US
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: 3+ Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: WordPerfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/442,629
FILING DATE: 18-NOV-6291219-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/913,805A
FILING DATE: 7 JAN 1998
APPLICATION NUMBER: PCT/JP97/00171
FILING DATE: 23 JAN 1997
APPLICATION NUMBER: JP 192260
FILING DATE: 22 JUL 1996
APPLICATION NUMBER: JP 162813
FILING DATE: 24 JUN 1996
APPLICATION NUMBER: JP 161648
FILING DATE: 21 JUN 1996
APPLICATION NUMBER: JP 10365
FILING DATE: 24 JAN 1996
ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 2356/3
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-429-1776
TELEFAX: 202-429-0796
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

; LENGTH: 575 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-442-629-2

Query Match 63.3%; Score 31; DB 4; Length 575;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
; : |||||
Db 379 IEEYTVHV 386

RESULT 22
US-08-982-785A-2
; Sequence 2, Application US/08982785A
; Patent No. 6258929
; GENERAL INFORMATION:
; APPLICANT: Kosik, Kenneth S.
; APPLICANT: Zhou, Jianhua
; TITLE OF INVENTION: ALARM RELATED PEPTIDES AND
; NUCLEIC ACIDS AND DIAGNOSIS USING THEM
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/982,785A
; FILING DATE: 02-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,556
; FILING DATE: 02-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 05311/018001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/542-5070
; TELEFAX: 617/542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 756 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
US-08-982-785A-2

Query Match 63.3%; Score 31; DB 4; Length 756;
Best Local Similarity 71.4%; Pred. No. 2.1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 3 NEYTVHL 9
; : |||||
Db 714 SEYTMHL 720

RESULT 23
US-08-447-031A-2
; Sequence 2, Application US/08447031A

; Patent No. 5851794
; GENERAL INFORMATION:
; APPLICANT: GUSS, Bengt
; APPLICANT: HOOK, Magnus
; APPLICANT: JONSSON, Hans
; APPLICANT: LINDBERG, Martin
; APPLICANT: PATTI, Joseph
; APPLICANT: SIGNAS, Christer
; APPLICANT: SWITALSKI, Lech
; TITLE OF INVENTION: A COLLAGEN BINDING PROTEIN AS WELL AS
; ITS PREPARATION
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Burns, Doane, Swecker & Mathis
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/447,031A
; FILING DATE: 22-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/861,804
; FILING DATE: 21-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/SE91/00707
; FILING DATE: 22-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9003374-7
; FILING DATE: 22-OCT-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: McGowan, Malcolm K.
; REGISTRATION NUMBER: 39,300
; REFERENCE/DOCKET NUMBER: 012889-006
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1183 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-447-031A-2

Query Match 63.3%; Score 31; DB 2; Length 1183;
Best Local Similarity 55.6%; Pred. No. 3.4e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
; : |||||
Db 1069 KVGNYTTHV 1077

RESULT 24
PCT-US94-05150-12
; Sequence 12, Application PC/TUS9405150
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Purified Mammalian Fit3 Ligands and Agonists and Antagonist
; NUMBER OF SEQUENCES: 37
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5


```
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05150
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/162,413
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,111
; FILING DATE: 19-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,391
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,340
; FILING DATE: 13-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/092,549
; FILING DATE: 16-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,263
; FILING DATE: 07-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,231
; FILING DATE: 19-MAY-1993
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 28 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US94-05150-12

Query Match 61.2%; Score 30; DB 5; Length 28;
Best Local Similarity 66.7%; Pred. No. 8.8;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
| | | | |
Db 18 KPRLTVHL 26

RESULT 25
PCT-US94-05150-17
; Sequence 17, Application PC/TUS9405150
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Purified Mammalian Flt3 Ligands and Agonists and Antagonists
; NUMBER OF SEQUENCES: 37
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Macintosh
; OPERATING SYSTEM: Macintosh 6.0.5
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/05150
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/162,413
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/155,111
; FILING DATE: 19-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,391
; FILING DATE: 24-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/106,340
; FILING DATE: 13-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/092,549
; FILING DATE: 16-JUL-1993
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/089,263
; FILING DATE: 07-JUL-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/065,231
; FILING DATE: 19-MAY-1993
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
PCT-US94-05150-17

Query Match 61.2%; Score 30; DB 5; Length 42;
Best Local Similarity 66.7%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
| | | | |
Db 18 KPRLTVHL 26

RESULT 26
US-08-900-927-1
; Sequence 1, Application US/08900927
; Patent No. 5840537
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,927
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0350 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 593 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: MMLR20T01
; CLONE: 475485
US-08-900-927-1

Query Match 61.2%; Score 30; DB 2; Length 593;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
```

QY 1 KMNEYTVHL 9
:|:|: ||
Db 339 ELNKYSTHL 347

RESULT 27
US-08-900-927-3
; Sequence 3, Application US/08900927
; Patent No. 5840537
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,927
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0350 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 593 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 642026
US-08-900-927-3

Query Match 61.2%; Score 30; DB 2; Length 593;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
:|:|: ||
Db 339 ELNKYSTHL 347

RESULT 28
US-08-900-927-4
; Sequence 4, Application US/08900927
; Patent No. 5840537
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.

QY 1 KMNEYTVHL 9
:|:|: ||
Db 339 ELNKYSTHL 347

RESULT 29
US-09-191-279-1
; Sequence 1, Application US/09191279
; Patent No. 5981192
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191,279
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,927
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0350 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 593 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 1246217
US-08-900-927-4

Query Match 61.2%; Score 30; DB 2; Length 593;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
:|:|: ||
Db 339 ELNKYSTHL 347

RESULT 30
US-09-191-279-1
; Sequence 1, Application US/09191279
; Patent No. 5981192
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Corley, Neil C.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/191,279
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,927
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749

REFERENCE/DOCKET NUMBER: PF-0350 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 593 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: MMLR20T01
CLONE: 475485
US-09-191-279-1

Query Match 61.2%; Score 30; DB 2; Length 593;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 339 ELNKYSTHL 347

RESULT 30
US-09-191-279-3
Sequence 3, Application US/09191279
Patent No. 5981192
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Corley, Neil C.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191.279
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/900.927
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0350 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 593 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 642026
US-09-191-279-3

Query Match 61.2%; Score 30; DB 2; Length 593;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 339 ELNKYSTHL 347

RESULT 31
US-09-191-279-4
Sequence 4, Application US/09191279
Patent No. 5981192
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Corley, Neil C.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/191.279
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/900.927
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0350 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 593 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1246217
US-09-191-279-4

Query Match 61.2%; Score 30; DB 2; Length 593;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 339 ELNKYSTHL 347

RESULT 32
US-09-334-476-1
Sequence 1, Application US/09334476
Patent No. 6162901
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Corley, Neil C.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.

STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/334,476
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/191,279
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0350 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 593 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: MMLR20T01
CLONE: 475485
US-09-334-476-1

Query Match 61.2%; Score 30; DB 4; Length 593;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
::|:|:|

DB 339 ELNKYSTHL 347

RESULT 33
US-09-334-476-3
Sequence 3, Application US/09334476
Patent No. 6162901
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Corley, Neil C.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/334,476
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/191,279
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0350 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 593 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 642026
US-09-334-476-3

Query Match 61.2%; Score 30; DB 4; Length 593;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
::|:|:|

DB 339 ELNKYSTHL 347

RESULT 34
US-09-334-476-4
Sequence 4, Application US/09334476
Patent No. 6162901
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Corley, Neil C.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NEW VESICLE TRANSPORT PROTEIN
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/334,476
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/191,279
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0350 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 593 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 1246217
US-09-334-476-4

Query Match 61.2%; Score 30; DB 4; Length 593;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;

Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db 339 ELNKYSTHL 347

RESULT 35
US-09-012-030-2
; Sequence 2, Application US/09012030
; Patent No. 5912169
; GENERAL INFORMATION:
; APPLICANT: SCHMIDT, Ralf-Michael, STITT, Marc, SONNEWALD,
; APPLICANT: Uwe
; TITLE OF INVENTION: Transketolase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1
; CURRENT APPLICATION DATA:
; FILING DATE: US/09/012,030
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/590,454
; FILING DATE: 22-JAN-1996
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 743 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-012-030-2

Query Match 61.2%; Score 30; DB 2; Length 743;
Best Local Similarity 62.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
Db 335 KANSYSVH 342

RESULT 36
US-08-590-454-2
; Sequence 2, Application US/08590454
; Patent No. 5925535
; GENERAL INFORMATION:
; APPLICANT: SCHMIDT, Ralf-Michael, STITT, Marc, SONNEWALD,
; APPLICANT: Uwe
; TITLE OF INVENTION: Transketolase
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kell & Weinkauff
; STREET: 1101 Connecticut Avenue
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage
; COMPUTER: IBM AT-compatible, 80486 processor
; OPERATING SYSTEM: MS-DOS version 6.0
; SOFTWARE: WordPerfect version 5.1

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/590,454
; FILING DATE: 22-JAN-1996
; CLASSIFICATION: 435
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 743 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-590-454-2

Query Match 61.2%; Score 30; DB 2; Length 743;
Best Local Similarity 62.5%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
Db 335 KANSYSVH 342

RESULT 37
US-09-074-579-4
; Sequence 4, Application US/09074579
; Patent No. 6001596
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPsin-TYPE
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/074,579
; FILING DATE: HEREWITH
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0505 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 911 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENEABK
; CLONE: gi33989
US-09-074-579-4

Query Match 61.2%; Score 30; DB 3; Length 911;
Best Local Similarity 50.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
| :|:|
Db 143 MEQFTIHL 150

RESULT 38
US-09-388-774-4
; Sequence 4, Application US/09388774
; Patent No. 6228991
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Guegler, Karl J.
; APPLICANT: Patterson, Chandra
; TITLE OF INVENTION: GROWTH-ASSOCIATED TRYPSIN-TYPE
; TITLE OF INVENTION: INHIBITOR HEAVY CHAIN PRECURSOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Word Perfect 6.1/MS-DOS 6.2
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/388-774
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/074,579
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Cerrone, Michael C
; REGISTRATION NUMBER: 39,132
; REFERENCE/DOCKET NUMBER: PF-0505 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 911 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GENEABK
; CLONE: g133989
US-09-388-774-4

Query Match 61.2%; Score 30; DB 4; Length 911;
Best Local Similarity 50.0%; Pred. No. 4.1e+02;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
| :|:|
Db 143 MEQFTIHL 150

RESULT 39
US-08-449-645A-18
; Sequence 18, Application US/08449645A
; Patent No. 5981245
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine
; TITLE OF INVENTION: Kinases
; NUMBER OF SEQUENCES: 43

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,645A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-449-645A-18

Query Match 61.2%; Score 30; DB 2; Length 976;
Best Local Similarity 62.5%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
| :|:|
Db 917 KMQYTEH 924

RESULT 40
US-08-702-367A-18
; Sequence 18, Application US/08702367A
; Patent No. 5981246
; GENERAL INFORMATION:
; APPLICANT: Fox, Gary M.
; TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine
; TITLE OF INVENTION: Kinases
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Amgen Patent Operations/RBW
; STREET: 1840 Dehavilland Drive
; CITY: Thousand Oaks
; STATE: California
; COUNTRY: USA
; ZIP: 91320
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/702,367A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Winter, Robert B.
; REFERENCE/DOCKET NUMBER: A-287
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 976 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-702-367A-18

Query Match 61.2%; Score 30; DB 2; Length 976;
Best Local Similarity 62.5%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNEYTVH 8
DB 917 KMOQYTEH 924

RESULT 41
PCT-US95-04681-18
Sequence 18, Application PC/TUS9504681
GENERAL INFORMATION:
APPLICANT: Fox, Gary M.
TITLE OF INVENTION: Eph-Like Receptor Protein Tyrosine Kinases
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Patent Operations/RBW
STREET: 1840 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04681
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Winter, Robert B.
REFERENCE/DOCKET NUMBER: A-287
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 976 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-04681-18

Query Match 61.2%; Score 30; DB 5; Length 976;
Best Local Similarity 62.5%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNEYTVH 8
DB 917 KMOQYTEH 924

RESULT 42
US-08-673-789-8
Sequence 8, Application US/08673789
Patent No. 5814479
GENERAL INFORMATION:
APPLICANT: ZHOU, RENPING; SCHULZ, NICHOLAS,
APPLICANT: T.; KROMER, LAWRENCE, F.; VANDE WOUDE,
APPLICANT: GEORGE, F.
TITLE OF INVENTION: BSK RECEPTOR LIKE
TITLE OF INVENTION: TYROSINE KINASE AND LIGAND AND THEIR
TITLE OF INVENTION: USE IN DIAGNOSTIC AND THERAPEUTIC
METHODS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORGAN & FINNEGAN
STREET: 345 PARK AVENUE
CITY: NEW YORK
STATE: NEW YORK

COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: FLOPPY DISK
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/673,789
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/177,812
FILING DATE: 04-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: CAROL M. GRUPPI
REGISTRATION NUMBER: 37,341
REFERENCE/DOCKET NUMBER: 2026-4105
TELEPHONE: (212) 751-4800
TELEFAX: (212) 751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 977
TYPE: AMINO ACID
STRANDEDNESS: UNKNOWN
TOPOLOGY: UNKNOWN
US-08-673-789-8

Query Match 61.2%; Score 30; DB 2; Length 977;
Best Local Similarity 62.5%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNEYTVH 8
DB 918 KMOQYTEH 925

RESULT 43
US-08-185-432-19
Sequence 19, Application US/08185432
Patent No. 5750652
GENERAL INFORMATION:
APPLICANT: Artavanis-Tsakonas, Spyridon
APPLICANT: Busseau, Isabelle
APPLICANT: Diederich, Robert J.
APPLICANT: Xu, Tian
APPLICANT: Matsuno, Kenji
TITLE OF INVENTION: DELTEX PROTEINS, NUCLEIC ACIDS, AND
TITLE OF INVENTION: ANTIBODIES, AND RELATED METHODS AND COMPOSITIONS
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: PENNIE & EDMONDS
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/185,432
FILING DATE: 21-JAN-1994
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: MISTOCK, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7326-006
TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 2703 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-185-432-19

Query Match 61.2%; Score 30; DB 1; Length 2703;
Best Local Similarity 83.3%; Pred. No. 1.4e+03;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 3 NEYTVH 8
IIIIII
Db 1080 NEYTVH 1085

RESULT 44
US-08-534-910B-6
Sequence 6, Application US/08534910B
Patent No. 5766911

GENERAL INFORMATION:
APPLICANT: KOIKE, Ayumi
APPLICANT: OBATA, Shusei
APPLICANT: NISHINO, Tokuzo
APPLICANT: OHNUMA, Shinichi
APPLICANT: NAKAZAWA, Takeshi
APPLICANT: OGURA, Kyoza
APPLICANT: KOYAMA, Tanetoshi
TITLE OF INVENTION: Mutated Farnesylidiphosphate Synthase Capable
OF SYNTHESIZING GERANYLGERANYLDIPHOSPHATE AND GENE CODING THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: U.S.
ZIP: 20036-5405

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.25" Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: IBM/Word Perfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/534,910B
FILING DATE: 28-SEPT-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-25253
FILING DATE: 14-FEB-1995

ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 77670/398
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)429-1776
TELEFAX: (202)429-0796
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacillus stearothermophilus
US-08-534-910B-6

Query Match 59.2%; Score 29; DB 1; Length 297;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
II :I:II
Db 186 KMLOYSVH 193

RESULT 45

US-08-534-910B-7
Sequence 7, Application US/08534910B
Patent No. 5766911
GENERAL INFORMATION:
APPLICANT: KOIKE, Ayumi
APPLICANT: OBATA, Shusei
APPLICANT: NISHINO, Tokuzo
APPLICANT: OHNUMA, Shinichi
APPLICANT: NAKAZAWA, Takeshi
APPLICANT: OGURA, Kyoza
APPLICANT: KOYAMA, Tanetoshi
TITLE OF INVENTION: Mutated Farnesylidiphosphate Synthase Capable
OF SYNTHESIZING GERANYLGERANYLDIPHOSPHATE AND GENE CODING THEREOF
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kenyon & Kenyon
STREET: 1025 Connecticut Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: U.S.
ZIP: 20036-5405

COMPUTER READABLE FORM:
MEDIUM TYPE: 3.25" Floppy Disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS 6.2
SOFTWARE: IBM/Word Perfect 6.1 Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/534,910B
FILING DATE: 28-SEPT-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-25253
FILING DATE: 14-FEB-1995

ATTORNEY/AGENT INFORMATION:
NAME: Toffenetti, Judith L.
REGISTRATION NUMBER: 39,048
REFERENCE/DOCKET NUMBER: 77670/398
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)429-1776
TELEFAX: (202)429-0796
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 297 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Bacillus stearothermophilus
US-08-534-910B-7

Query Match 59.2%; Score 29; DB 1; Length 297;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
II :I:II
Db 186 KMLOYSVH 193

Search completed: November 6, 2002, 12:06:20
Job time : 10.1111 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 5.33333 Seconds
(without alignments)
65.339 Million cell updates/sec

Title: US-09-905-083-31
Perfect score: 49
Sequence: 1 KMNEYTVHL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	49	100.0	253	1	KLK7_HUMAN
2	39	79.6	218	1	VLXS_BPK3
3	39	79.6	218	1	VLXS_BPT4
4	38	77.6	261	1	RFA4_HUMAN
5	34	69.4	266	1	RFA2_RAT
6	34	69.4	270	1	RFA2_HUMAN
7	34	69.4	270	1	RFA2_MOUSE
8	34	69.4	285	1	Y539_AQUAE
9	33	67.3	204	1	FIXJ_RHIME
10	33	67.3	218	1	FCA2_TRYBB
11	33	67.3	229	1	FCA4_TRYBB
12	33	67.3	233	1	FCAL_TRYBB
13	33	67.3	261	1	TFPB_MORBO
14	33	67.3	407	1	FCA3_TRYBB
15	33	67.3	429	1	SECY_AQUAE
16	33	67.3	799	1	RS64_MOUSE
17	33	67.3	804	1	RS64_HUMAN
18	33	67.3	2185	1	PYRL_DICDI
19	32	65.3	116	1	RL31_CHLRE
20	32	65.3	491	1	CAD3_BOVIN
21	32	65.3	500	1	XYNB_THESA
22	32	65.3	500	1	XYNB_THESJ
23	32	65.3	639	1	TET5_ENTFA
24	32	65.3	639	1	TET5_STRPN
25	32	65.3	688	1	ARK2_HUMAN
26	32	65.3	752	1	CLPE_STRPN
27	32	65.3	987	1	YD94_METJA
28	32	65.3	1127	1	Y855_TREPA
29	31	63.3	306	1	TRUB_NEIMA
30	31	63.3	307	1	TRUB_NEIMB
31	31	63.3	639	1	TET1_ENTFA
32	31	63.3	639	1	TET9_ENTFA
33	31	63.3	639	1	TETM_NEIME

34	31	63.3	639	1	TETM_STAAU	Q53770 staphylococ
35	31	63.3	639	1	TETM_UREUR	P09757 ureaplasma
36	31	63.3	818	1	PTK2_YEAST	P47116 saccharomyc
37	31	63.3	870	1	COPG_CAEEL	Q22498 caenorhabdl
38	31	63.3	1183	1	CNA_STAAU	Q53654 staphylococ
39	30	61.2	134	1	TSHB_CHICK	O57340 gallus gall
40	30	61.2	276	1	SACT_BACSU	P26212 bacillus su
41	30	61.2	296	1	FTR_METKA	O49610 methanopyru
42	30	61.2	320	1	K6PL_VIBCH	O9knp2 vibrio chol
43	30	61.2	407	1	PE10_CAEEL	P41991 caenorhabdl
44	30	61.2	445	1	NSMA_CAEEL	O45870 caenorhabdl
45	30	61.2	464	1	MURF_BORBU	Q44777 borrelia bu

ALIGNMENTS

RESULT 1
KLK7_HUMAN
ID KLK7_HUMAN STANDARD; PRT; 253 AA.
AC P49862:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic
DE enzyme) (hSCCE).
GN KLK7 OR PRSS6 OR SCCE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.
RX TISSUE=Skin;
RC MEDLINE=94308225; PubMed=8034709;
RA Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A.,
RA Egelrud T.;
RT "Cloning, expression, and characterization of stratum corneum
RT chymotryptic enzyme. A skin-specific human serine proteinase.";
RL J. Biol. Chem. 269:19420-19426(1994).
RN [2]
SEQUENCE FROM N.A.
RC TISSUE=Keratinocytes;
RA Yousef G.M., Scorrilas A., Diamandis E.P.;
RT "Molecular characterization, mapping and tissue expression of the
RT human stratum corneum chymotryptic enzyme gene.";
RN [3]
SEQUENCE FROM N.A.
RC MEDLINE=20510030; PubMed=11054574;
RA Can L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
RA Moss P., Paepel B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene
RT cluster located in chromosome 19q13 region.";
RN [4]
SEQUENCE FROM N.A.
RC MEDLINE=95314630; PubMed=7794273;
RA Skjott A., Stroemqvist M., Egelrud T.;
RT "Primary substrate specificity of recombinant human stratum corneum
RT chymotryptic enzyme.";
RN [5]
FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE
STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS
SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID
RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE
CLEAVES INSULIN B CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE-
1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION
OF PRECURSORS TO INFLAMMATORY CYTOKINES.
CC -!
TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS
EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE
ALSO SEEN IN THE BRAIN AND KIDNEY.
CC -!
SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L33404; AAC37551.1; -;
DR EMBL: AF166330; AAD49718.1; -;
DR EMBL: AF243527; AAG33360.1; -;
DR HSSP: P00763; LDPO.
DR MEROPS: S01.300; -;
DR MIM: 604438; -;
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; zymogen; Glycoprotein; Signal.
FT SIGNAL 1 22
FT PROPEP 23 29 ACTIVATION PEPTIDE.
FT CHAIN 30 253 KALLIKREIN 7.
FT ACT_SITE 70 70 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 36 137 BY SIMILARITY.
FT DISULFID 55 71 BY SIMILARITY.
FT DISULFID 137 239 BY SIMILARITY.
FT DISULFID 144 211 BY SIMILARITY.
FT DISULFID 176 190 BY SIMILARITY.
FT DISULFID 201 226 BY SIMILARITY.
FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 253 AA; 27525 MW; 2D68B6B15A76A668 CRC64;

Query Match 100.0%; Score 49; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.0073;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMEYTVHL 9
IIIIIIII
DB 72 KMEYTVHL 80

RESULT 2
VLYS_BPK3
ID VLYS_BPK3 STANDARD; PRT; 218 AA.
AC P10393.
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last annotation update)
DE Lysis protein.
GN T.
OS Bacteriophage K3.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae.
OX NCBI_TaxID=10674;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87250254; PubMed=3597316;
RA Riede I.;
RT "Lysis gene t of T-even bacteriophages: evidence that colicins and
bacteriophage genes have common ancestors.";
RL J. Bacteriol. 169:2956-2961(1987).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Y00408; CAA68470.1; -;

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: M16812; AAA88415.1; -;
DR PIR: A27083; YVBPk3.
KW Phage lysis protein.
SQ SEQUENCE 218 AA; 25222 MW; 21B4DC02ACA0ECF6 CRC64;

Query Match 79.6%; Score 39; DB 1; Length 218;
Best Local Similarity 87.5%; Pred. No. 0.7;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
I:IIIIII
DB 139 MDEYTVHL 146

RESULT 3
VLYS_BPT4
ID VLYS_BPT4 STANDARD; PRT; 218 AA.
AC P06808.
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Lysis protein (Holin) (Protein rv).
GN T OR RV.
OS Bacteriophage T4.
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
OC T4-like phages.
OX NCBI_TaxID=10665;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87316934; PubMed=3628006;
RA Montag D., Degen M., Henning U.;
RT "Nucleotide sequence of gene t (lysis gene) of the E. coli phage T4.";
RL Nucleic Acids Res. 15:6736-6736(1987).
RN [2]
RP SEQUENCE FROM N.A.
RA Kutter E., Arisaka F., Kunisawa T., Tsugita A., Mosig G.,
Mesyazhinov V., Ruger W., Stidham T., Thomas E.;
RT "Bacteriophage T4 genome analysis.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 1-34 FROM N.A.
RX MEDLINE=88011316; PubMed=2958637;
RA Montag D., Riede I., Eschbach M.-L., Degen M., Henning U.;
RT "Receptor-recognizing proteins of T-even type bacteriophages.
constant and hypervariable regions and an unusual case of
evolution.";
RL J. Mol. Biol. 196:165-174(1987).
RN [4]
RP SEQUENCE OF 202-218 FROM N.A.
RC STRAIN=D;
RX MEDLINE=93106978; PubMed=8416914;
RA Orsini G., Ouhammouch M., le Caer J.P., Brody E.N.;
RT "The asia gene of bacteriophage T4 codes for the anti-sigma 70
protein.";
RL J. Bacteriol. 175:85-93(1993).
CC -!- FUNCTION: AT THE END OF THE GROWTH CYCLE, PHAGE T4 EXPRESSES TWO
CC GENES WITH LYSIS FUNCTION, E AND T. NOTHING IS KNOWN ABOUT THE
CC GENE PRODUCT OF T, ALTHOUGH IT HAS BEEN SUGGESTED THAT IT ACTS AS
CC A PHOSPHOLIPASE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: Y00408; CAA68470.1; -;

```
DR EMBL; AF158101; AAD42661.1; -.
DR EMBL; X05677; CAA29164.1; -.
DR EMBL; M94441; AAA32481.1; -.
DR PIR; JF0028; YVBP4.
KW Phage lysis protein; Transmembrane.
FT TRANSMEM 35 49 POTENTIAL.
SQ SEQUENCE 218 AA; 25175 MW; 9110BE11D772DF5 CRC64;

Query Match 79.6%; Score 39; DB 1; Length 218;
Best Local Similarity 87.5%; Pred. No. 0.7;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
DB 139 MDEYTVHL 146
I:|||||

RESULT 4
RFA4_HUMAN STANDARD; PRT; 261 AA.
AC Q13156;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Replication protein A 30 kDa subunit (RP-A) (RF-A) (Replication
DE factor-A protein 4).
GN RPA4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95280910; PubMed=7760808;
RA Keshav K.F., Chen C., Dutta A.;
RT "Rpa4, a homolog of the 34-kilodalton subunit of the replication
RT protein A complex.";
RL Mol. Cell. Biol. 15:3119-3128(1995).
RN [2]
RP REVISION TO 10.
RA Keshav K.F.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ABSOLUTELY REQUIRED FOR SIMIAN VIRUS 40 DNA REPLICATION
CC IN VITRO. IT PARTICIPATES IN A VERY EARLY STEP IN INITIATION. RP-A
CC IS A SINGLE-STRANDED DNA-BINDING PROTEIN (BY SIMILARITY).
CC -!- SUBUNIT: HETEROPRIMER OF 70, 32/30, AND 14 kDa CHAINS. THE DNA-
CC BINDING ACTIVITY MAY RESIDE EXCLUSIVELY ON THE 70 kDa SUBUNIT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN PLACENTAL AND
CC COLON MUCOSA.
CC -!- SIMILARITY: TO OTHER SPECIES RFA2/RPA2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U24186; AAB08488.2; -.
DR InterPro; IPR002309; trna_synt_2.
DR Pfam; PF01336; trna_ant1; 1.
KW DNA replication; Nuclear protein; Alternative splicing.
SQ SEQUENCE 261 AA; 28868 MW; 6A925FAEDBE21718 CRC64;

Query Match 77.6%; Score 38; DB 1; Length 261;
Best Local Similarity 75.0%; Pred. No. 1.4;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
DB 151 MNEFTVHI 158
I:|||||
```

```
RESULT 5
RFA2_RAT STANDARD; PRT; 266 AA.
ID Q63528;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Replication protein A 32 kDa subunit (RP-A) (RF-A) (Replication
DE factor-A protein 2) (Fragment).
GN RPA2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Nagelhus T., Haug T., Krokan H.E.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ABSOLUTELY REQUIRED FOR SIMIAN VIRUS 40 DNA REPLICATION
CC IN VITRO. IT PARTICIPATES IN A VERY EARLY STEP IN INITIATION. RP-A
CC IS A SINGLE-STRANDED DNA-BINDING PROTEIN.
CC -!- SUBUNIT: HETEROPRIMER OF 70, 32, AND 14 kDa CHAINS. THE
CC DNA-BINDING ACTIVITY MAY RESIDE EXCLUSIVELY ON THE 70 kDa SUBUNIT.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- PTM: PHOSPHORYLATED IN A CELL-CYCLE-DEPENDENT MANNER (FROM THE S
CC PHASE UNTIL MITOSIS) (BY SIMILARITY).
CC -!- SIMILARITY: TO OTHER SPECIES RFA2/RPA2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X98490; CAA67116.1; -.
DR InterPro; IPR002309; trna_synt_2.
DR Pfam; PF01336; trna_ant1; 1.
KW DNA replication; Phosphorylation; Nuclear protein.
SQ SEQUENCE 266 AA; 28827 MW; 52E13704CID2E13A CRC64;

Query Match 69.4%; Score 34; DB 1; Length 266;
Best Local Similarity 62.5%; Pred. No. 9.2;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
DB 148 MNEFTAHI 155
I:|||||

RESULT 6
RFA2_HUMAN STANDARD; PRT; 270 AA.
ID P15927;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Replication protein A 32 kDa subunit (RP-A) (RF-A) (Replication
DE factor-A protein 2).
GN RPA2 OR RPA32 OR RPA2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=90153966; PubMed=2406247;
RA Erdile L.F., Wold M.S., Kelly T.J.;
RT "The primary structure of the 32-kDa subunit of human replication
```

RT protein A.";
RL J. Biol. Chem. 265:3177-3182(1990).
CC -|- FUNCTION: ABSOLUTELY REQUIRED FOR SIMIAN VIRUS 40 DNA REPLICATION
CC IN VITRO. IT PARTICIPATES IN A VERY EARLY STEP IN INITIATION. RP-A
CC IS A SINGLE-STRANDED DNA-BINDING PROTEIN.
CC -|- SUBUNIT: HETEROPRIMER OF 70, 32/30, AND 14 kDa CHAINS. THE DNA-
CC BINDING ACTIVITY MAY RESIDE EXCLUSIVELY ON THE 70 kDa SUBUNIT.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- PTM: PHOSPHORYLATED IN A CELL-CYCLE-DEPENDENT MANNER (FROM THE S
CC PHASE UNTIL MITOSIS).
CC -|- SIMILARITY: TO OTHER SPECIES RFA2/RPA2.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J05249; AAA36560.1; -
CC DR PIR: A43711; A43711.
CC DR MIM: 179836;
CC DR InterPro: IPR002309; trna-synt_2.
CC DR Pfam: PF01336; trna-anti; 1.
CC KW DNA replication; Phosphorylation; Nuclear protein.
CC FT DOMAIN 1 29 GLY/SER-RICH.
CC FT DOMAIN 37 45 ARG/LYS-RICH (BASIC).
CC FT DOMAIN 95 123 ASP/GLU-RICH (ACIDIC).
CC FT DOMAIN 127 145 ARG/LYS-RICH (BASIC).
CC FT DOMAIN 247 270 ASP/GLU-RICH (ACIDIC).
CC SQ SEQUENCE 270 AA; 29247 MW; 61A563EA7B34A9B1 CRC64;
Query Match 69.4%; Score 34; DB 1; Length 270;
Best Local Similarity 62.5%; Pred. No. 9.4;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 MNEYTVHL 9
DB 152 MNEFTTHI 159
RESULT 7
ID RFA2_MOUSE STANDARD; PROT: 270 AA.
AC Q62193;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Replication protein A 32 kDa subunit (RP-A) (RP-A) (Replication
DE factor-A protein 2).
GN RPA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91334146; PubMed=1908076;
RA Nakagawa M., Tsukada S., Soma T., Shimizu Y.K., Miyake S.,
RA Iwakatsu A., Sugiyama H.;
RT "cDNA cloning of the murine 30-kDa protein homologous to the 32-kDa
RT subunit of human replication protein A.";
RL Nucleic Acids Res. 19:4292-4292(1991).
CC -|- FUNCTION: ABSOLUTELY REQUIRED FOR SIMIAN VIRUS 40 DNA REPLICATION
CC IS A SINGLE-STRANDED DNA-BINDING PROTEIN.
CC IN VITRO. IT PARTICIPATES IN A VERY EARLY STEP IN INITIATION. RP-A
CC -|- SUBUNIT: HETEROPRIMER OF 70, 32, AND 14 kDa CHAINS. THE
CC DNA-BINDING ACTIVITY MAY RESIDE EXCLUSIVELY ON THE 70 kDa SUBUNIT.
CC -|- SUBCELLULAR LOCATION: Nuclear.
CC -|- PTM: PHOSPHORYLATED IN A CELL-CYCLE-DEPENDENT MANNER (FROM THE S
CC PHASE UNTIL MITOSIS) (BY SIMILARITY).
CC -|- SIMILARITY: TO OTHER SPECIES RFA2/RPA2.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D00812; BAA00693.1; -
CC DR MGD: MGI:1339939; Rpa2.
CC DR InterPro: IPR002309; trna-synt_2.
CC DR Pfam: PF01336; trna-anti; 1.
CC KW DNA replication; Phosphorylation; Nuclear protein.
CC SQ SEQUENCE 270 AA; 29718 MW; BF0EF86612A48011 CRC64;
Query Match 69.4%; Score 34; DB 1; Length 270;
Best Local Similarity 62.5%; Pred. No. 9.4;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 MNEYTVHL 9
DB 152 MNEFTTHI 159
RESULT 8
ID Y539_AQUAE STANDARD; PROT: 285 AA.
AC Q66818;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein AQ_539.
GN AQ_539.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aufay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus".
RL Nature 392:353-358(1998).
CC -|- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE000694; AAC06778.1; -
CC DR Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 7 29 POTENTIAL.
FT TRANSMEM 49 71 POTENTIAL.
FT TRANSMEM 95 117 POTENTIAL.
FT TRANSMEM 137 156 POTENTIAL.
FT TRANSMEM 232 254 POTENTIAL.
FT TRANSMEM 259 281 POTENTIAL.
SQ SEQUENCE 285 AA; 33543 MW; FDC54632F0EF99BB CRC64;
Query Match 69.4%; Score 34; DB 1; Length 285;
Best Local Similarity 66.7%; Pred. No. 9.9;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMNEYTVHL 9
DB 152 MNEFTTHI 159

```

Db 207 KFNEYTLNL 215
RESULT 9
FIXJ_RHIME STANDARD; PRT; 204 AA.
AC PI0958;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcriptional regulatory protein fixJ.
GN FIXJ OR RA0669 OR SMA1227.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OG Plasmid pSymA (megaplasmid 1).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88311059; PubMed=2842062;
RA David M., Davenport M.-L., Batut J., Dedieu A., Domergue O., Ghai J.,
RA Hertig C., Bolstad P., Kahn D.;
RT "Cascade regulation of nif gene expression in Rhizobium meliloti.";
RN Cell 54:671-683(1988).
RL [2]
RP SEQUENCE FROM N.A.
RX STRAIN=1021.
RA Barnett M.J., Fisher R.F., Jones T., Komp C., Abola A.P.,
RA Barloy-Hubier F., Bowser L., Capela D., Galibert F., Gouzy J.,
RA Gurjal M., Hong A., Huizar L., Hyman R.W., Kahn D., Kahn M.L.,
RA Kallman S., Keating D.H., Palm C., Peck M.C., Surzycki R., Wells D.H.,
RA Yeh K.-C., Davis R.W., Federspiel N.A., Long S.R.;
RT "Nucleotide sequence and predicted functions of the entire
RT Sinorhizobium meliloti pSymA megaplasmid.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9883-9888(2001).
RN [3]
RP SEQUENCE OF 200-204 FROM N.A.
RX MEDLINE=89305532; PubMed=2663474;
RA Batut J., Davenport-Mingot M.-L., David M., Jacobs J., Garnerone A.-M.,
RA Kahn D.;
RT "fixK, a gene homologous with fnr and crp from Escherichia coli,
RT regulates nitrogen fixation genes both positively and negatively in
RT Rhizobium meliloti.";
RL EMBO J. 8:1279-1286(1989).
RN [4]
RP REVIEW, AND MUTAGENESIS OF HTH REGION.
RX MEDLINE=91312138; PubMed=1857213;
RA Kahn D., Ditta G.S.;
RT "Modular structure of FixJ: homology of the transcriptional activator
RT domain with the -35 binding domain of sigma factors.";
RL Mol. Microbiol. 5:987-997(1991).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS) OF 1-126.
RX MEDLINE=201313479; PubMed=10647182;
RA Gouet P., Fabry B., Guillet V., Birck C., Mourey L., Kahn D.,
RA Samama J.-P.;
RT "Structural transitions in the FixJ receiver domain.";
RL Structure 7:1517-1526(1999).
CC -!- FUNCTION: FIXJ, WHEN ACTIVATED BY FIXL, INDUCES THE EXPRESSION OF
CC BOTH NIFA, REQUIRED FOR ACTIVATION OF CLASSICAL NIF AND FIX GENES,
CC AND FIXK, REQUIRED FOR FIXN ACTIVATION.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -!- PTM: PHOSPHORYLATED BY FIXL.
CC -!- SIMILARITY: CONTAINS 1 RESPONSE REGULATORY DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE LUXR/UHPA FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
-----
entties requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: J03174; AAA26276.1; -
CC EMBL: Z21854; CAN79898.1; -
CC EMBL: AE007255; AAK65327.1; -
CC EMBL: X15079; CAA33182.1; -
CC PIR: B31227; B31227.
CC PIR: S39985; S39985.
CC PIR: S32838; S32838.
CC PDB: IDB; 26-NOV-99.
CC PDB: IDCK; 26-NOV-99.
CC InterPro: IPR00792; HTH_LuxR.
CC InterPro: IPR001789; Response_reg.
CC Pfam: PF00196; GerE; 1.
CC Pfam: PF00072; response_reg; 1.
CC PRINTS: PR00038; HTHLUXR.
CC SMART: SM00421; HTH_LUXR; 1.
CC SMART: SM00448; REC; 1.
CC PROSITE: PS00622; HTH_LUXR_FAMILY; 1.
CC PROSITE: PS0110; RESPONSE_REGULATORY; 1.
CC Sensory transduction; Phosphorylation; Transcription regulation;
KW DNA-binding; Activator; Nitrogen fixation; Plasmid; 3D-structure;
KW Complete proteome.
FT DOMAIN 1 119 RESPONSE REGULATORY.
FT MOD_RES 54 54 PHOSPHORYLATION.
FT DNA_BIND 159 178 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 204 AA; 22218 MW; 2EDA356967352292 CRC64;
Query Match 67.3%; Score 33; DB 1; Length 204;
Best Local Similarity 62.5%; Pred. No. 11;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 MNEYTVHL 9
DB 1 MTDYTVHI 8
RESULT 10
FCA2_TRYBB STANDARD; PRT; 218 AA.
ID FCA2_TRYBB AC Q26680;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Flagellar calcium-binding protein TB-24 (24 kDa calflagin) (24 kDa
DE calclmedin).
OS Trypanosoma brucei brucei.
OC Eukaryota; Eulenczoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=427;
RX MEDLINE=95118301; PubMed=7818488;
RA Wu Y., Deford J., Benjamin R., Lee M.G.-S., Ruben L.;
RT "The gene family of EF-hand calcium-binding proteins from the
RT flagellum of Trypanosoma brucei.";
RL Biochem. J. 304:833-841(1994).
CC -!- FUNCTION: MAY CONTRIBUTE TO THE RAPID MOTILITY OF THE
CC TRYPANOSOMES, PLAYING A ROLE EITHER IN FLAGELLAR STRUCTURE OR IN
CC CALCIUM METABOLISM. COULD ALTERNATE BETWEEN A GDP-BOUND INACTIVE
CC FORM TO A CALCIUM/GTP-BOUND ACTIVE FORM.
CC -!- SUBCELLULAR LOCATION: FLAGELLUM.
CC -!- DOMAIN: THIS PROTEIN HAS FOUR EF-HAND DOMAINS, THREE OF WHICH MAY
CC BE FUNCTIONAL CALCIUM-BINDING SITES.
CC -!- SIMILARITY: BELONGS TO THE EF-HAND CALCIUM BINDING PROTEIN
CC SUPERFAMILY; CALFLAGIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; U06644; AAB40004.1; -
 DR HSSP; P02590; LAP4.
 DR InterPro; IPR003299; Calflagin.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00036; ehand; 3.
 DR PRINTS; PR01362; CALFLAGIN.
 DR SMART; SM00054; EFh; 3.
 DR PROSITE; PS00018; EF_HAND; 2.
 KW Calcium-binding; Repeat; Flagella.
 FT CA_BIND 61 72
 FT CA_BIND 110 121 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 121 ANCESTRAL CALCIUM SITE 2 (POTENTIAL).
 FT CA_BIND 143 154 EF-HAND 3 (POTENTIAL).
 FT CA_BIND 180 191 EF-HAND 4 (POTENTIAL).
 SQ SEQUENCE 218 AA; 24252 MW; 041152AC2BDD6DDB CRC64;

Query Match 67.3%; Score 33; DB 1; Length 218;
 Best Local Similarity 55.6%; Pred. No. 12;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMEYTVHL 9
 DB 82 KLDEFTTHL 90

RESULT 11
 FCA4_TRYBB
 ID FCA4_TRYBB STANDARD; PRT; 229 AA.
 AC Q26677;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Flagellar calcium-binding protein TB-1.7G (17 kDa calflagin) (17 kDa
 DE calclmedin) (fragment).
 OS Trypanosoma brucei brucei.
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5702;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=427;
 RX MEDLINE=95118301; PubMed=7818488;
 RA Wu Y., Deford J., Benjamin R., Lee M.G.-S., Ruben L.;
 RT "The gene family of EF-hand calcium-binding proteins from the
 RL flagellum of Trypanosoma brucei";
 RL Biochem. J. 304:833-841(1994).
 CC -!- FUNCTION: MAY CONTRIBUTE TO THE RAPID MOTILITY OF THE
 CC TRYPANOSOMES, PLAYING A ROLE EITHER IN FLAGELLAR STRUCTURE OR IN
 CC CALCIUM METABOLISM. COULD ALTERNATE BETWEEN A GDP-BOUND INACTIVE
 CC FORM TO A CALCIUM/GTP-BOUND ACTIVE FORM.
 CC -!- SUBCELLULAR LOCATION: FLAGELLUM.
 CC -!- DOMAIN: THIS PROTEIN HAS FOUR EF-HAND DOMAINS, THREE OF WHICH MAY
 CC BE FUNCTIONAL CALCIUM-BINDING SITES.
 CC -!- SIMILARITY: BELONGS TO THE EF-HAND CALCIUM BINDING PROTEIN
 CC SUPERFAMILY; CALFLAGIN FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; U05882; AAY75502.1; -
 DR HSSP; P02590; LAP4.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00036; ehand; 3.
 DR SMART; SM00054; EFh; 3.
 DR PROSITE; PS00018; EF_HAND; 2.
 KW Calcium-binding; Repeat; Flagella.
 FT NON_TER 1 1

FT CA_BIND 57 68 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 106 117 ANCESTRAL CALCIUM SITE 2 (POTENTIAL).
 FT CA_BIND 139 150 EF-HAND 3 (POTENTIAL).
 FT CA_BIND 176 187 EF-HAND 4 (POTENTIAL).
 SQ SEQUENCE 229 AA; 25109 MW; 675A1607E390743D CRC64;

Query Match 67.3%; Score 33; DB 1; Length 229;
 Best Local Similarity 55.6%; Pred. No. 13;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMEYTVHL 9
 DB 78 KLDEFTTHL 86

RESULT 12
 FCA4_TRYBB
 ID FCA4_TRYBB STANDARD; PRT; 233 AA.
 AC P17882;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Flagellar calcium-binding protein TB-17 (17 kDa calflagin) (17 kDa
 DE calclmedin).
 DE GN FCABP.
 OS Trypanosoma brucei brucei.
 OC Eukaryota; Euzlenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5702;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=427;
 RX MEDLINE=90332427; PubMed=2198539;
 RA Lee M.G.-S., Chen J., Ho A.W.M., D'Alessandro P., van der Ploeg L.H.T.;
 RT "A putative flagellar Ca2(+)-binding protein of the flagellum of
 RT trypanosomatid protozoan parasites";
 RL Nucleic Acids Res. 18:4252-4252(1990).
 CC -!- FUNCTION: MAY CONTRIBUTE TO THE RAPID MOTILITY OF THE
 CC TRYPANOSOMES, PLAYING A ROLE EITHER IN FLAGELLAR STRUCTURE OR IN
 CC CALCIUM METABOLISM. COULD ALTERNATE BETWEEN A GDP-BOUND INACTIVE
 CC FORM TO A CALCIUM/GTP-BOUND ACTIVE FORM.
 CC -!- SUBCELLULAR LOCATION: FLAGELLUM.
 CC -!- DOMAIN: THIS PROTEIN HAS FOUR EF-HAND DOMAINS, THREE OF WHICH MAY
 CC BE FUNCTIONAL CALCIUM-BINDING SITES.
 CC -!- SIMILARITY: BELONGS TO THE EF-HAND CALCIUM BINDING PROTEIN
 CC SUPERFAMILY; CALFLAGIN FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 CC EMBL; X53464; CAA37558.1; -
 DR PIR; S10515; AQUT17.
 DR HSSP; P02633; 2BCB.
 DR InterPro; IPR003299; Calflagin.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00036; ehand; 3.
 DR PRINTS; PR01362; CALFLAGIN.
 DR SMART; SM00054; EFh; 3.
 DR PROSITE; PS00018; EF_HAND; 2.
 KW Calcium-binding; Repeat; Flagella.
 FT CA_BIND 61 72 EF-HAND 1 (POTENTIAL).
 FT DOMAIN 110 121 ANCESTRAL CALCIUM SITE 2 (POTENTIAL).
 FT CA_BIND 143 154 EF-HAND 3 (POTENTIAL).
 FT CA_BIND 180 191 EF-HAND 4 (POTENTIAL).
 SQ SEQUENCE 233 AA; 25492 MW; 2ACF56FA38D08FC9 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 233;
 Best Local Similarity 55.6%; Pred. No. 13;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

CC STRAIN-VF5;
 RX MEDLINE-98196666; PubMed=9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus";
 RL Nature 392:353-358(1998).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SECA AND SECE
 CC TO ALLOW THE TRANSLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE,
 CC BY FORMING PART OF A CHANNEL.
 CC -1- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT
 CC COMPRISE THE PROKARYOTIC PROTEIN TRANSLOCATION APPARATUS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE SECY/SEC61-ALPHA FAMILY.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF060672; AAC06435.1; -
 DR InterPro; IPR002208; SecY.
 DR Pfam: PF00344; secy; 1.
 DR PRINTS: PR00303; SECYTRNLCASE.
 DR PROSITE: PS00755; SECY_1; 1.
 DR PROSITE: PS00756; SECY_2; 1.
 KW Protein transport; Translocation; Transmembrane; Complete proteome.
 FT TRANSMEM 19 39 POTENTIAL.
 FT TRANSMEM 72 92 POTENTIAL.
 FT TRANSMEM 115 135 POTENTIAL.
 FT TRANSMEM 152 172 POTENTIAL.
 FT TRANSMEM 184 204 POTENTIAL.
 FT TRANSMEM 213 233 POTENTIAL.
 FT TRANSMEM 272 292 POTENTIAL.
 FT TRANSMEM 309 329 POTENTIAL.
 FT TRANSMEM 367 387 POTENTIAL.
 FT TRANSMEM 388 408 POTENTIAL.
 SQ SEQUENCE 429 AA; 48100 MW; 87D4FC133D037C25 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 429;
 Best Local Similarity 66.7%; Pred. No. 25;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KNEYTVHL 9
 |:||||:|
 DB 111 KNEYTKYL 119

RESULT 16
 RSG4_MOUSE
 ID RSG4_MOUSE STANDARD; PRT; 799 AA.
 AC Q92268;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE RASGAP-activating-like protein 1.
 GN RASAL1 OR RASAL.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98426153; PubMed=9751798;
 RA Allen M., Chu S., Brill S., Stotler C., Buckler A.;
 RT "Restricted tissue expression pattern of a novel human rasGAP-related
 RT gene and its murine ortholog.";
 RL Gene 218:17-25(1998).
 CC -1- FUNCTION: PROBABLE INHIBITORY REGULATOR OF THE RAS-CYCLIC AMP

CC PATHWAY.
 CC -1- SIMILARITY: CONTAINS 2 C2 DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 BTK DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 RAS-GAP DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AF086714; AAD09007.1; -
 DR HSP; P04410; 1A25.
 DR MGD; MGI:1330842; Rasal1.
 DR InterPro; IPR001562; BTK.
 DR InterPro; IPR000008; C2.
 DR InterPro; IPR001849; PH.
 DR InterPro; IPR001936; RasGAP.
 DR Pfam; PF00779; BTK; 1.
 DR Pfam; PF00168; C2; 2.
 DR Pfam; PF00169; PH; 1.
 DR Pfam; PF00616; RasGAP; 1.
 DR SMART; SM00107; BTK; 1.
 DR SMART; SM00239; C2; 2.
 DR SMART; SM00233; PH; 1.
 DR SMART; SM00323; RasGAP; 1.
 DR PROSITE; PS00003; PH_DOMAIN; 1.
 DR PROSITE; PS00499; C2_DOMAIN_1; 2.
 DR PROSITE; PS00004; C2_DOMAIN_2; 2.
 DR PROSITE; PS00509; RAS_GTPASE_ACTIV_1; FALSE_NEG.
 DR PROSITE; PS00018; RAS_GTPASE_ACTIV_2; 1.
 KW GTPase activation; Repeat.
 FT DOMAIN 1 88 C2 DOMAIN 1.
 FT DOMAIN 135 216 C2 DOMAIN 2.
 FT DOMAIN 300 510 RAS-GAP.
 FT DOMAIN 565 672 PH.
 FT DOMAIN 674 710 BTK.
 FT DOMAIN 498 501 POLY-LEU.
 SQ SEQUENCE 799 AA; 89428 MW; E42F54B677F52269 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 799;
 Best Local Similarity 100.0%; Pred. No. 50;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9
 |:|||||
 DB 56 EYTVHL 61

RESULT 17
 RSG4_HUMAN
 ID RSG4_HUMAN STANDARD; PRT; 804 AA.
 AC O95294;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE RASGAP-activating-like protein 1.
 GN RASAL1 OR RASAL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-98426153; PubMed=9751798;
 RA Allen M., Chu S., Brill S., Stotler C., Buckler A.;
 RT "Restricted tissue expression pattern of a novel human rasGAP-related
 RT gene and its murine ortholog.";
 RL Gene 218:17-25(1998).
 RN [2]

RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21154917; PubMed=11230166;
RA Wleemann S., Well B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansorge W., Boecher M., Blocker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoef A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.-W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a catalog of human genes and proteins: sequencing and
analysis of 500 novel complete protein coding human cDNAs.";
RL Genome Res. 11:422-435(2001).
CC -!- FUNCTION: PROBABLE INHIBITORY REGULATOR OF THE RAS-CYCLIC AMP
PATHWAY.
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN THYROID AND ADRENAL
MEDULLA, LOWER EXPRESSION IN BRAIN, SPINAL CORD AND TRACHEA.
CC -!- SIMILARITY: CONTAINS 2 C2 DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 BTK DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 RAS-GAP DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF086713; AD090006.1; -;
DR EMBL; ALI36672; CAB66607.1; -;
DR HSSP; P04410; 1A25.
DR MIM; 604118; -;
DR InterPro; IPR001562; BTK.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR001936; RASGAP.
DR Pfam; PF00779; BTK; 1.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00616; RASGAP; 1.
DR SMART; SM00107; BTK; 1.
DR SMART; SM00239; C2; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00323; RASGAP; 1.
DR PROSITE; PS50003; PH_DOMAIN_1; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; 2.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
DR PROSITE; PS00509; RAS_GTPASE_ACTIV_1; FALSE_NEG.
DR PROSITE; PS50018; RAS_GTPASE_ACTIV_2; 1.
KW GTPase activation; Repeat.
FT DOMAIN 1 88 C2 DOMAIN 1.
FT DOMAIN 135 216 C2 DOMAIN 2.
FT DOMAIN 301 511 RAS-GAP.
FT DOMAIN 565 672 PH.
FT DOMAIN 674 710 BTK.
FT DOMAIN 499 502 POLY-LEU.
FT CONFLICT 141 141 H -> Q (IN REF. 2).
FT CONFLICT 321 321 H -> R (IN REF. 2).
FT CONFLICT 346 346 L -> P (IN REF. 2).
FT CONFLICT 514 514 Q -> R (IN REF. 2).
FT CONFLICT 552 552 E -> EE (IN REF. 2).
FT CONFLICT 701 804 AAGCSRTHSAVTLGDSWDLDPDAEAQTVYRQLLLGRDQLR
FT LKLLDSNMDDTLLEADTGACPEVLARQRAATARLLVLAIDL
FT DRAHEEFQOQERGAALGLGP -> GEGAGRTAGSALPTR
FT QGAGVPLPGLAVQVPWASYSTSLCLLLNCMKRLITVPISR
FT SCDN (IN REF. 2).
SQ SEQUENCE 804 AA; 90005 MW; 0379A72358002872 CRC64;
Query Match 67.3%; Score 33; DB 1; Length 804;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9

DB 56 EYTVHL 61
|||||
RESULT 18
PYRI_DICDI
ID PYRI_DICDI STANDARD; PRT; 2185 AA.
AC P20054;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein PYRI-3 [Includes: Glutamine-dependent carbamoyl-phosphate
DE synthase (EC 6.3.5.5); Aspartate carbamoyltransferase (EC 2.1.3.2);
DE Dihydroorotase (EC 3.5.2.3)].
GN PYRI-3.
OS Dictyostellum discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE OF 1-467 AND 1175-2185 FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=89137111; PubMed=2917570;
RA Faure M., Camonis J.H., Jacquet M.;
RT "Molecular characterization of a Dictyostellum discoideum gene
RT encoding a multifunctional enzyme of the pyrimidine pathway.";
RL Eur. J. Biochem. 179:345-358(1989).
RN [2]
RP SEQUENCE OF 367-1408 FROM N.A.
RC STRAIN=AX2;
RX MEDLINE=92329976; PubMed=1627825;
RA Elgar G., Schofield J.P.;
RT "Carbamoyl phosphate synthetase (CPSase) in the PYRI-3 multigene of
RT Dictyostellum discoideum.";
RL DNA Seq. 2:219-226(1992).
CC -!- FUNCTION: THIS PROTEIN IS A "FUSION" PROTEIN ENCODING FOUR
CC ENZYMIC ACTIVITIES OF THE PYRIMIDINE PATHWAY (GATASE, CPSASE,
CC ATCASE AND DHOASE).
CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
CC phosphate + glutamate + carbamoyl phosphate.
CC -!- CATALYTIC ACTIVITY: Carbamoyl phosphate + L-aspartate = phosphate
CC + N-carbamoyl-L-aspartate.
CC -!- CATALYTIC ACTIVITY: (S)-dihydroorotate + H(2)O = N-carbamoyl-L-
CC aspartate.
CC -!- COFACTOR: DHOASE ACTIVITY REQUIRES A ZINC ATOM.
CC -!- ENZYME REGULATION: ALLOSTERICALLY REGULATED AND CONTROLLED BY
CC PHOSPHORYLATION. PRPP IS AN ACTIVATOR WHILE UMP IS AN INHIBITOR
CC OF THE CPSASE REACTION.
CC -!- PATHWAY: FIRST TO THIRD STEP IN PYRIMIDINE BIOSYNTHESIS.
CC -!- SUBUNIT: HOMOHXAMER.
CC -!- SUBCELLULAR LOCATION: SEEN DURING GROWTH BUT NOT DURING
CC DEVELOPMENT.
CC -!- MISCELLANEOUS: GATASE (GLUTAMINE AMIDOTRANSFERASE) AND CPSASE
CC (CARBAMOYL PHOSPHATE SYNTHASE) FORM TOGETHER THE
CC GLUTAMINE-DEPENDENT CPSASE (GD-CPSASE) (EC 6.3.5.5).
CC -!- SIMILARITY: THE CPSASE DOMAIN IS SIMILAR TO OTHER CPSASES.
CC -!- SIMILARITY: IN THE CENTRAL SECTION, BELONGS TO THE DHOASE FAMILY.
CC -!- SIMILARITY: THE GATASE DOMAIN BELONGS TO TYPE-1 GLUTAMINE
CC AMIDOTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14633; CAA32781.1; -;
DR EMBL; X14634; CAA32782.1; -;
DR EMBL; X55433; CAA39077.1; -;
DR PIR; S02800; OZDOP3.
DR PIR; S23738; S23738.

```
DR HSP; P00479; 3CSU.
DR DictyDb; DD05019; pyr1-3.
DR InterPro; IPR001317; CPS_GATase.
DR InterPro; IPR000901; CPSase.
DR InterPro; IPR002474; CPSase_sm_chain.
DR InterPro; IPR002029; Carbmyltransf_asor.
DR InterPro; IPR002195; Dihydroorotase.
DR InterPro; IPR000991; GATase_1.
DR InterPro; IPR000384; MGS.
DR Pfam; PF00289; CPSase_L_chain; 3.
DR Pfam; PF02786; CPSase_L_D2; 3.
DR Pfam; PF02787; CPSase_L_D3; 1.
DR Pfam; PF00988; CPSase_sm_chain; 1.
DR Pfam; PF00744; Dihydroorotase; 1.
DR Pfam; PF00117; GATase; 1.
DR Pfam; PF00185; ORCace; 1.
DR Pfam; PF02729; ORCace_N; 1.
DR PRINTS; PR00100; AOTCASE.
DR PRINTS; PR00098; CPSASE.
DR PRINTS; PR00099; CPSGATASE.
DR PRINTS; PR00096; GATASE.
DR PROSITE; PS00097; CARBAMOYLTRANSFERASE; 1.
DR PROSITE; PS00442; GATASE_TYPE_1; 1.
DR PROSITE; PS00482; DIHYDROOROTASE_1; 1.
DR PROSITE; PS00483; DIHYDROOROTASE_2; 1.
DR PROSITE; PS00866; CPSASE_1; 2.
DR PROSITE; PS00867; CPSASE_2; 2.
KW Pyrimidine biosynthesis; Ligase; Transferase; Hydrolase; Zinc;
KW Multifunctional enzyme.
FT DOMAIN 1 351 GATASE (GLUTAMINE AMIDOTRANSFERASE).
FT DOMAIN 352 366 LINKER.
FT DOMAIN 367 1422 CPSASE (CARBAMOYL-PHOSPHATE SYNTHASE).
FT DOMAIN 367 909 CPSASE A.
FT DOMAIN 910 1422 CPSASE B.
FT DOMAIN 1423 1757 DHOASE (DIHYDROOROTASE).
FT DOMAIN 1758 1876 LINKER.
FT DOMAIN 1877 2185 ATCASE (ASPARTATE TRANSCARBAMYLASE).
FT ACT_SITE 236 236 GATASE (BY SIMILARITY).
FT ACT_SITE 322 322 GATASE (BY SIMILARITY).
FT ACT_SITE 324 324 GATASE (BY SIMILARITY).
FT METAL 1439 1439 ZINC (POTENTIAL).
FT METAL 1441 1441 ZINC (POTENTIAL).
FT METAL 1185 1192 NNEIKVIE -> TMKSKLSN (IN REF. 1).
FT CONFLICT 1362 1362 L -> LF (IN REF. 1).
SQ SEQUENCE 2185 AA; 241741 MW; D0F32D879F06F7E5 CRC64;

Query Match 67.3%; Score 33; DB 1; Length 2185;
Best Local Similarity 66.7%; Pred. NO. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEVTVHL 9
DB 1354 KMTENTFHL 1362

RESULT 19
RL31_CHLRE
ID RL31_CHLRE STANDARD; PRT; 116 AA.
AC P45841;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE 60S ribosomal protein L31.
GN RPL31.
OS Chlamydomonas reinhardtii.
OC Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
OC Chlamydomonadaceae; Chlamydomonas.
OX NCBI_TaxID=3055;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=137c / CC=125;
RA Dumont F., Joris B., Gunusboga A., Bruyninx M., Loppes R.;

RT "Isolation and characterization of cDNA sequences controlled by
RT inorganic phosphate in Chlamydomonas reinhardtii.";
RL Plant Sci. 89:55-67(1993).
CC -|- SIMILARITY: BELONGS TO THE L31E FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X66413; CAA47044.1; -.
CC InterPro; IPR000054; Ribosomal_L31e.
CC Pfam; PF01198; Ribosomal_L31e; 1.
CC ProDom; PD006030; Ribosomal_L31e; 1.
CC PROSITE; PS01144; RIBOSOMAL_L31E; 1.
KW Ribosomal protein.
SQ SEQUENCE 116 AA; 13254 MW; 1F4303FFCBF1FA6E CRC64;

Query Match 65.3%; Score 32; DB 1; Length 116;
Best Local Similarity 83.3%; Pred. NO. 9.5;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9
DB 15 EYTIHL 20

RESULT 20
CAD3_BOVIN
ID CAD3_BOVIN STANDARD; PRT; 491 AA.
AC P19535;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cadherin-3 (Placental-cadherin) (P-cadherin) (Fragment).
GN CDH3 OR CDHP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90360979; PubMed=2390969;
RA Liaw C.W., Cannon C., Power M.D., Kiboneka P.K., Rubin L.L.;
RT "Identification and cloning of two species of cadherins in bovine
RT endothelial cells.";
RL EMBO J. 9:2701-2708(1990).
CC -|- FUNCTION: CADHERINS ARE CALCIUM DEPENDENT CELL ADHESION PROTEINS.
CC THEY PREFERENTIALLY INTERACT WITH THEMSELVES IN A HOMOPHILIC
CC MANNER IN CONNECTING CELLS; CADHERINS MAY THUS CONTRIBUTE TO THE
CC SORTING OF HETEROGENEOUS CELL TYPES.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein.
CC -|- SIMILARITY: CONTAINS 5 CADHERIN DOMAINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X53614; CAA37676.1; -.
CC PIR; S11694; IJBOCP.
CC HSSP; P15116; INCJ.
DR InterPro; IPR002126; Cadherin.
DR InterPro; IPR000233; Cadherin_C_term.
DR Pfam; PF00028; cadherin; 3.
DR Pfam; PF01049; Cadherin_C_term; 1.
```

DR SMART; SM00112; CA; 2.
DR PROSITE; PS00232; CADHERIN_1; 1.
DR PROSITE; PS0268; CADHERIN_2; 2.
KW Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat.
FT NON_TER 1
FT DOMAIN <1 316 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 317 339 POTENTIAL.
FT DOMAIN 340 491 CYTOPLASMIC (POTENTIAL).
FT DOMAIN <1 102 CADHERIN 3.
FT DOMAIN 103 208 CADHERIN 4.
FT DOMAIN 209 314 CADHERIN 5.
FT DOMAIN 447 462 SER-RICH.
FT CARBOHYD 228 228 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 491 AA; 54207 MW; 6BE80AE5918C4771 CRC64;

Query Match 65.3%; Score 32; DB 1; Length 491;
Best Local Similarity 66.7%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
| |||||
DB 273 KQGEYDVHL 281

RESULT 21
XYNB_THESA STANDARD; PRT; 500 AA.
AC P36906;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Beta-xylosidase (EC 3.2.1.37) (1,4-beta-D-xylan xylohydrolase)
DE (Xylan 1,4-beta-xylosidase).
GN XYNB.
OS Thermoanaerobacter saccharolyticum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Thermoanaerobacterium.
OX NCBI_TaxID=28896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6A-RI;
RX MEDLINE=93367395; PubMed=8360617;
RA Lee Y.E., Zeikus J.G.;
RT "Genetic organization, sequence and biochemical characterization of recombinant beta-xylosidase from Thermoanaerobacterium saccharolyticum strain B6A-RI";
RT J. Gen. Microbiol. 139:1235-1243 (1993).
RN [2]
RP CHARACTERIZATION.
RC STRAIN=B6A-RI;
RX MEDLINE=96195684; PubMed=8612648;
RA Armand S., Vielle C., Gey C., Heyraud A., Zeikus J.G., Henriessat B.;
RT "Stereochemical course and reaction products of the action of beta-xylosidase from Thermoanaerobacterium saccharolyticum strain B6A-RI";
RT Eur. J. Biochem. 236:706-713 (1996).
RN [3]
RP ACTIVE SITE GLU-277.
RX MEDLINE=98437142; PubMed=9761746;
RA Vocadlo D.J., Mackenzie L.F., He S., Zeikus G.J., Withers S.G.;
RT "Identification of Glu-277 as the catalytic nucleophile of Thermoanaerobacterium saccharolyticum beta-xylosidase using electrospray MS";
RT Biochem. J. 335:449-455 (1998).
RL
CC -1- FUNCTION: HAS HYDROLYTIC ACTIVITY TOWARDS XYLOPENTOSE, XYLOTRIOSE, CC XYLOBIOSIDE AND P-NITROPHENYL-BETA-D-XYLOPYRANOSIDE, BUT HAS NO CC ACTIVITY TOWARD XYLAN. OPTIMAL TEMPERATURE IS 70 DEGREES CELSIUS;
CC LOOSE ACTIVITY AT 85 DEGREES CELSIUS.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-XYLANS SO AS TO CC REMOVE SUCCESSIVE D-XYLOSE RESIDUES FROM THE NON-REDUCING CC TERMINI. IT ALSO HYDROLYSES XYLOBIOSIDE.
CC -1- PATHWAY: XYLAN DEGRADATION.
CC -1- INDUCTION: BY XYLAN AND XYLOSE.

CC -1- MASS SPECTROMETRY: MW=58666; MW_ERR=6; METHOD=Electrospray.
CC -1- SIMILARITY: BELONGS TO FAMILY 39 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M97883; AAA27369.1; -
DR PIR; S41859; S41859.
DR InterPro; IPR000514; Glyco_hydro_39.
DR Pfam; PF01229; Glyco_hydro_39; 1.
DR PRINTS; PR00745; GLHYDRLASE39.
DR PROSITE; PS01027; GLYCOSYL_HYDROL_F39; 1.
KW Xylan degradation; Hydrolase; Glycosidase.
FT ACT_SITE 160 160 PROTON DONOR (POTENTIAL).
FT ACT_SITE 277 277 NUCLEOPHILE.
SQ SEQUENCE 500 AA; 58606 MW; 3B6D59E70A5F4CBC CRC64;

Query Match 65.3%; Score 32; DB 1; Length 500;
Best Local Similarity 66.7%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
| |||||
DB 233 KQGEYTPHL 241

RESULT 22
XYNB_THESJ STANDARD; PRT; 500 AA.
ID XYNB_THESJ
AC O30360;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Beta-xylosidase (EC 3.2.1.37) (1,4-beta-D-xylan xylohydrolase)
DE (Xylan 1,4-beta-xylosidase).
GN XYNB OR XYLX.
OS Thermoanaerobacterium sp. (strain JW/SL YS485).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Thermoanaerobacter group; Thermoanaerobacterium.
OX NCBI_TaxID=60708;
RN [1]
RP SEQUENCE FROM N.A.
RA Lorenz W.W., Wiegel J.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF 1,4-BETA-D-XYLANS SO AS TO
CC REMOVE SUCCESSIVE D-XYLOSE RESIDUES FROM THE NON-REDUCING
CC TERMINI. IT ALSO HYDROLYSES XYLOBIOSIDE.
CC -1- SIMILARITY: BELONGS TO FAMILY 39 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF001926; AAB68820.1; -
DR InterPro; IPR000514; Glyco_hydro_39.
DR Pfam; PF01229; Glyco_hydro_39; 1.
DR PRINTS; PR00745; GLHYDRLASE39.
DR PROSITE; PS01027; GLYCOSYL_HYDROL_F39; 1.
KW Xylan degradation; Hydrolase; Glycosidase.
FT ACT_SITE 160 160 PROTON DONOR (POTENTIAL).
FT ACT_SITE 277 277 NUCLEOPHILE (BY SIMILARITY).
SQ SEQUENCE 500 AA; 58483 MW; D194E6B76B89E10C CRC64;

Query Match 65.3%; Score 32; DB 1; Length 500;

Best Local Similarity 66.7%; Pred. No. 47;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
| ||| ||
Db 233 KQGEYTPHL 241

RESULT 23

TET5_ENTFA STANDARD; PRT; 639 AA;
AC P11131;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tetracycline resistance protein tetM from transposon Tn1545
DE (TetM(1545)).
GN TETM OR TET(M).
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BM4127; TRANSPOSON=Tn1545;
RX MEDLINE=87016342; PubMed=3020504;
RA Martin P., Trieu-Cuot P., Courvalin P.;
RT "Nucleotide sequence of the tetM tetracycline resistance determinant
of the streptococcal conjugative shuttle transposon Tn1545.";
RL Nucleic Acids Res. 14:7047-7058(1986).
CC -!- FUNCTION: ABOLISH THE INHIBITORY EFFECT OF TETRACYCLIN ON PROTEIN
SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC TETM/TETO SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC EMBL; X04388; CAA27977.1; -;
CC PIR; A24333; A24333.
CC HSP; P13551; 1FNM.
CC InterPro: IPR000640; EFG_C.
CC InterPro: IPR000795; GTP_EFTU.
CC Pfam: PF00679; EFG_C; 1.
CC Pfam: PF00009; GTP_EFTU; 1.
CC Pfam: PF03144; GTP_EFTU_D2; 1.
CC PRINTS; PR00315; ELONGATNCT.
CC PROSITE; PS00301; EFATOR_GTP; 1.
CC Protein biosynthesis; Antibiotic resistance; GTP-binding;
KW Transposable element.
FT NP_BIND 10 17 GTP (BY SIMILARITY).
FT NP_BIND 74 78 GTP (BY SIMILARITY).
FT NP_BIND 128 131 GTP (BY SIMILARITY).
SQ SEQUENCE 639 AA; 72657 MW; 2F7A3CD0588253CE CRC64;

Query Match 65.3%; Score 32; DB 1; Length 639;
Best Local Similarity 55.6%; Pred. No. 62;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
| ||| ||
Db 426 KKAETIHI 434

RESULT 24

TETM_STRPN STANDARD; PRT; 639 AA.
ID TETM_STRPN

AC

DT

DT

DT

DE

GN

OS

OC

OX

RN

RP

RC

RX

RA

RT

RL

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Q54807;

15-DEC-1998 (Rel. 37, Created)

15-DEC-1998 (Rel. 37, Last sequence update)

15-DEC-1998 (Rel. 37, Last annotation update)

Tetracycline resistance protein tetM from transposon Tn5251 (Tet(M)).

TETM(5251).

Streptococcus pneumoniae.

Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;

Streptococcus.

NCBI_TaxID=1313;

[1]

SEQUENCE FROM N.A.

STRAIN-DP 1322;

MEDLINE=96162866; PubMed=8595862;

Provvedi R., Manganelli R., Pozzi G.;

"Characterization of conjugative transposon Tn5251 of Streptococcus

pneumoniae.";

FEMS Microbiol. Lett. 135:231-236(1996).

-!- FUNCTION: ABOLISH THE INHIBITORY EFFECT OF TETRACYCLIN ON PROTEIN

SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.

-!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.

TETM/TETO SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL Outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>or send an email to license@isb-sib.ch).

EMBL; X90939; CAA62436.1; -;

HSP; P13551; 1FNM.

InterPro: IPR000640; EFG_C.

InterPro: IPR000795; GTP_EFTU.

Pfam: PF00679; EFG_C; 1.

Pfam: PF00009; GTP_EFTU; 1.

Pfam: PF03144; GTP_EFTU_D2; 1.

PRINTS; PR00315; ELONGATNCT.

PROSITE; PS00301; EFATOR_GTP; 1.

Protein biosynthesis; Antibiotic resistance; GTP-binding;

KW Transposable element.

FT NP_BIND 10 17 GTP (BY SIMILARITY).

FT NP_BIND 74 78 GTP (BY SIMILARITY).

FT NP_BIND 128 131 GTP (BY SIMILARITY).

SQ SEQUENCE 639 AA; 72556 MW; F18131E08B81F3C0 CRC64;

Query Match 65.3%; Score 32; DB 1; Length 639;

Best Local Similarity 55.6%; Pred. No. 62;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

| ||| ||

Db 426 KKAETIHI 434

RESULT 25

ARK2_HUMAN

ID ARK2_HUMAN STANDARD; PRT; 688 AA.

AC P35626; Q9UGW9;

DT 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Beta-adrenergic receptor kinase 2 (EC 2.7.1.126) (Beta-ARK-2)

DE (G-protein coupled receptor kinase 3).

GN ADRBK2 OR GRK3 OR BARK2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

[1]

RP SEQUENCE FROM N.A.

CC TISSUE=Brain;
RX MEDLINE=93151831; PubMed=8427589;
RA Parruti G., Ambrosini G., Sallesse M., de Blasi A.;
RT "Molecular cloning, functional expression and mRNA analysis of human
beta-adrenergic receptor kinase 2.";
RL Biochem. Biophys. Res. Commun. 190:475-481(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20057165; PubMed=10591208;
RA Dunham I., Hunt A.R., Collins J.E., Bruskewich R., Beare D.M.,
RA Clamp M., Smink L.J., Alnsough R., Almeida J.P., Babbage A.,
RA Bagguley C., Bailey J., Barlow K., Bates K.N., Beasley O., Bird C.P.,
RA Blakey S., Bridgeman A.M., Buck D., Burgess J., Burrill W.D.,
RA Burton J., Carder C., Carter N.P., Chen Y., Clark G., Clegg S.M.,
RA Cobley V., Cole C.G., Collier R.E., Connor R.E., Conroy D., Corby N.,
RA Coville G.J., Cox A.V., Davis J., Dawson E., Dhani P.D., Dockree C.,
RA Dodsworth S.J., Durbin R.M., Ellington A., Evans K.L., Fey J.M.,
RA Fleming K., French L., Garner A.A., Gilbert J.G.R., Goward M.E.,
RA Grahnam D., Griffiths M.N., Hall C., Hall R., Hall-Tamlyn G.,
RA Heathcott R.W., Ho S., Holmes S., Hunt S.E., Jones M.C., Kershaw J.,
RA Kimberley A., King A., Laird G.K., Langford C.F., Leversha M.A.,
RA Lloyd C., Lloyd D.M., Martyn I.D., Mashreghi-Mohammadi M.,
RA Matthews L., Mccann O.T., Mclay J., Mclaren S., Mennuray A.A.,
RA Milne S.A., Mortimore B.J., Odell C.N., Pavitt R., Pearce A.V.,
RA Pearson D., Phillips B.J., Phillips S.H., Plumb R.W., Ramsay H.,
RA Ramsey V., Rogers L., Ross M.T., Scott C.E., Sehra H.K., Skuce C.D.,
RA Smalley S., Smith M.L., Soderlund C., Spragon L., Steward C.A.,
RA Sulston J.E., Swann R.M., Vaudin M., Wall M., Wallis J.M.,
RA Whiteley M.N., Willey D., Williams L., Williams S., Williamson H.,
RA Wilmer T.E., Wilming L., Wright C.L., Hubbard T., Bentley D.R.,
RA Beck S., Rogers J., Shimizu N., Minoshima S., Kawasaki K., Sasaki T.,
RA Asakawa S., Kudoh J., Shintani A., Shibuya K., Yoshizaki Y., Aoki N.,
RA Mitsuyama S., Roe B.A., Chen F., Chu L., Crabtree J., Deschamps S.,
RA Do A., Dorman A., Fang F., Fu Y., Hu P., Hua A., Kenton S.,
RA Lai H., Lao H.I., Lewis S., Lewis S., Lin S.-P., Loh P., Mala J.E.,
RA Nguyen T., Pan H., Phan S., Qi S., Qian Y., Ray L., Ren Q., Shaull S.,
RA Sloan D., Song L., Wang Q., Wang Y., Wang Z., White J., Willingham D.,
RA Wu H., Yao Z., Zhan M., Zhang G., Chissoe S., Murray J., Miller N.,
RA Minx P., Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H.,
RA Bourne S., Cordes M., Du Z., Fulton L., Goela D., Graves T.,
RA Hawkins J., Hinds K., Kemp K., Latreille P., Layman D., Ozersky P.,
RA Rohlfing T., Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K.,
RA Nelson J., Korf I., Bedell J.A., Hillier L., Mardis E., Waterston R.,
RA Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA Edelman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA Khan A.S., Lane L., Tilahun Y., Wright H.;
RT "The DNA sequence of human chromosome 22.";
RL Nature 402:489-495(1999).
CC -|- FUNCTION: SPECIFICALLY PHOSPHORYLATES THE AGONIST-OCCUPIED FORM
CC OF THE BETA-ADRENERGIC AND CLOSELY RELATED RECEPTORS.
CC -|- CATALYTIC ACTIVITY: ATP + [beta-adrenergic receptor] = ADP +
CC [beta-adrenergic receptor] phosphate.
CC -|- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC GPRK SUBFAMILY.
CC -|- SIMILARITY: CONTAINS 1 PH DOMAIN.
CC -|- SIMILARITY: CONTAINS 1 RGS DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69117; CAB48870.1; -;
DR EMBL; AL022329; CAB45657.1; -;
DR HSSP; P25098; 1BAK.
DR MIM; 109636; -;
DR InterPro: IPR000719; Euk_pkinase.

DR InterPro: IPR000239; GPCR_kinase.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000961; pkinase_C.
DR InterPro: IPR000342; RGS.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00615; RGS; 1.
DR PRINTS; PR00717; GPCR_KINASE.
DR SMART; SM00233; PH; 1.
DR SMART; SM00315; RGS; 1.
DR SMART; SM00133; S_TK_X; 1.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00132; RGS; 1.
KW Transferase: Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 1 190 N-TERMINAL.
FT DOMAIN 191 453 PROTEIN_KINASE.
FT DOMAIN 454 688 C-TERMINAL.
FT DOMAIN 54 175 RGS.
FT DOMAIN 558 652 PH.
FT NP_BIND 197 205 ATP (BY SIMILARITY).
FT BINDING 220 220 ATP (BY SIMILARITY).
FT ACT_SITE 317 317 BY SIMILARITY.
FT CONFLICT 308 308 V -> M (IN REF. 2).
SQ SEQUENCE 688 AA; 79677 MW; 55D55E80B3FA0B73 CRC64;

Query Match 65.3%; Score 32; DB 1; Length 688;
Best Local Similarity 71.4%; Pred. No. 67;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVH 8
|||:|:
DB 188 MNEFSVH 194

RESULT 26
CLPE_STRPN STANDARD; PRT; 752 AA.
ID CLPE_STRPN
AC P35594;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent clp protease ATP-binding subunit clpE (Exported protein
DE 4).
GN CLPE OR EXP4 OR SP0820.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=21357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Maynam L.A., White O., Salzberg S.L., Lewis M.R., Radune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Utterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Angiuoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Lofcut B.J., Fang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.;
RT "Complete genome sequence of a virulent isolate of Streptococcus
RT pneumoniae".
RL Science 293:498-506(2001).
RN [2]
RP SEQUENCE OF 600-692 FROM N.A.
RC STRAIN=R6x;
RX MEDLINE=95020625; PubMed=7934910;
RA Pearce B.J., Yin Y.B., Masure H.R.;

```

RT "Genetic identification of exported proteins in Streptococcus
RL pneumoniae";
CC -1- FUNCTION: COULD BE NECESSARY FOR DEGRADING PROTEINS GENERATED BY
CC CERTAIN TYPES OF STRESS (By similarity).
CC -1- SUBCELLULAR LOCATION: Membrane-associated.
CC -1- INDUCTION: BY HEAT SHOCK (Probable).
CC -1- SIMILARITY: BELONGS TO THE CLPA/CLPB FAMILY. CLPE SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE007388; AAK74952.1; -.
CC TIGR: SP0820; -.
CC
CC InterPro: IPR003593; AAA.
CC InterPro: IPR003959; AAA-subfam.
CC InterPro: IPR001270; CLP-AB.
CC InterPro: IPR001943; UVR.
CC Pfam: PF00004; AAA; 2.
CC Pfam: PF02151; UVR; 1.
CC PRINTS: PR00300; CLP-PROTEASEA.
CC SMART: SM00382; AAA; 2.
CC PROSITE: PS00870; CLPAB_1; 1.
CC PROSITE: PS00871; CLPAB_2; 1.
CC Chapterone: ATP-binding; Heat shock; Zinc-finger; Repeat;
KW Complete proteome.
FT DOMAIN 113 360 I.
FT DOMAIN 413 604 II.
FT ZN_FING 3 32 C4-TYPE (POTENTIAL).
FT NP_BIND 158 165 ATP (POTENTIAL).
FT NP_BIND 487 494 ATP (POTENTIAL).
SQ SEQUENCE 752 AA; 83814 MW; 376989EA320F0C88 CRC64;

Query Match 65.3%; Score 32; DB 1; Length 752;
Best Local Similarity 55.6%; Pred. No. 74;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 7 KINDSTIHL 15

RESULT 27
YD94_METJA
ID YD94_METJA STANDARD; PRT; 987 AA.
AC Q58789;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MJ1394.
GN MJ1394.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kieruloff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;

```

```

RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RL jannaschii.";
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: TO M.JANNASCHII MJ1393 AND A.FULGIDUS AF2028.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67579; AAB99404.1; -.
CC TIGR: MJ1394; -.
CC
CC Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 958 978 POTENTIAL.
SQ SEQUENCE 987 AA; 112360 MW; D1E628FFB28CA86D CRC64;

Query Match 65.3%; Score 32; DB 1; Length 987;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 351 KMSEWTVNL 359

RESULT 28
Y855_TREPA
ID Y855_TREPA STANDARD; PRT; 1127 AA.
AC O83827;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0855 precursor.
GN TP0855.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RA Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA Dodson R., Gwinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artiach P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RL spirochete.";
RC Science 281:375-388(1998).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE001255; AAC65823.1; -.
CC TIGR: TP0855; -.
CC
CC Hypothetical protein; Signal; Coiled coil; Complete proteome.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 1127 HYPOTHETICAL PROTEIN TP0855.
FT DOMAIN 1076 1126 COILED COIL (POTENTIAL).
SQ SEQUENCE 1127 AA; 124699 MW; C10D4E622C484D39 CRC64;

```

```

Query Match      65.3%; Score 32; DB 1; Length 1127;
Best Local Similarity 55.6%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KNEYTVHL 9
DB 622 KRRYTAHM 630

RESULT 29
TRUB_NEIMA
ID TRUB_NEIMA STANDARD; PRT; 306 AA.
AC Q9JTX5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE tRNA pseudouridine synthase B (EC 4.2.1.70) (tRNA pseudouridine 55
DE synthase) (Psi155 synthase) (Pseudouridylate synthase) (Uracil
DE hydrolyase).
GN TRUB OR NMA1588.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli P., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moulé S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
CC -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL162756; CAB84815.1; -
DR InterPro: IPR002501; TruB_N.
DR Pfam: PF01509; TruB_N; 1.
KW Lyase; tRNA processing; Complete proteome.
SQ SEQUENCE 306 AA; 33515 MW; D066EDF862977FB CRC64;
-----
Query Match      63.3%; Score 31; DB 1; Length 306;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
DB 245 LNDYAVHM 252

RESULT 30
TRUB_NEIMB
ID TRUB_NEIMB STANDARD; PRT; 307 AA.
AC Q9JYV1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE tRNA pseudouridine synthase B (EC 4.2.1.70) (tRNA pseudouridine 55
DE synthase) (Psi155 synthase) (Pseudouridylate synthase) (Uracil
DE hydrolyase).
GN TRUB OR NMA1588.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli P., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Felwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moulé S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
CC -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL162756; CAB84815.1; -
DR InterPro: IPR002501; TruB_N.
DR Pfam: PF01509; TruB_N; 1.
KW Lyase; tRNA processing; Complete proteome.
SQ SEQUENCE 306 AA; 33515 MW; D066EDF862977FB CRC64;
-----
Query Match      63.3%; Score 31; DB 1; Length 307;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
DB 245 LNDYAVHM 253

RESULT 31
TETL_ENTFA
ID TETL_ENTFA STANDARD; PRT; 639 AA.
AC Q47810;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tetracycline resistance protein tetM from transposon TnF01 (Tet(M)).
GN TETM OR TET(M).
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F01; TRANSPOSON=TnF01;
RA Perreten V., Kolleff B., Teuber M.;
RT "Conjugal transfer of Tn916-like transposon TnF01 from
RT Enterococcus faecalis F01 to several Gram-positive bacteria."
RL Syst. Appl. Microbiol. 20:27-38(1997).
CC -!- FUNCTION: ABOLISH THE INHIBITORY EFFECT OF TETRACYCLIN ON PROTEIN
CC SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.
CC -----

```

```

DE synthase) (Psi155 synthase) (Pseudouridylate synthase) (Uracil
DE hydrolyase).
GN TRUB OR NMB1374.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Venathavan J.,
RA Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -!- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF002485; AAF41748.1; -
DR TIGR; NMB1374; -
DR InterPro: IPR002501; TruB_N.
DR Pfam: PF01509; TruB_N; 1.
KW Lyase; tRNA processing; Complete proteome.
SQ SEQUENCE 307 AA; 33632 MW; 4B0FBADACB999E79 CRC64;
-----
Query Match      63.3%; Score 31; DB 1; Length 307;
Best Local Similarity 50.0%; Pred. No. 44;
Matches 4; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
DB 246 LNDYAVHM 253

RESULT 31
TETL_ENTFA
ID TETL_ENTFA STANDARD; PRT; 639 AA.
AC Q47810;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tetracycline resistance protein tetM from transposon TnF01 (Tet(M)).
GN TETM OR TET(M).
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F01; TRANSPOSON=TnF01;
RA Perreten V., Kolleff B., Teuber M.;
RT "Conjugal transfer of Tn916-like transposon TnF01 from
RT Enterococcus faecalis F01 to several Gram-positive bacteria."
RL Syst. Appl. Microbiol. 20:27-38(1997).
CC -!- FUNCTION: ABOLISH THE INHIBITORY EFFECT OF TETRACYCLIN ON PROTEIN
CC SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.
CC -----

```

```
CC -|- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC TETM/TETO SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X92947; CAA63530.1; -.
DR HSSP; P13551; IFNM.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR000795; GTP_EFTU.
DR InterPro; IPR004161; GTP_EFTU_D2.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR PROSITE; PS00301; EFACOR_GTP; 1.
KW Protein biosynthesis; Antibiotic resistance; GTP-binding;
KW Transposable element.
FT NP_BIND 10 17 GTP (BY SIMILARITY).
FT NP_BIND 74 78 GTP (BY SIMILARITY).
FT NP_BIND 128 131 GTP (BY SIMILARITY).
SQ SEQUENCE 639 AA; 72588 MW; FAE14580EF8CE914 CRC64;

Query Match 63.3%; Score 31; DB 1; Length 639;
Best Local Similarity 55.6%; Pred. NO. 1e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 426 KNAEYTIHI 434
| | | | |
| | | | |

RESULT 32
TET9_ENTFA STANDARD; PRT; 639 AA.
ID TET9_ENTFA STANDARD; PRT; 639 AA.
AC P21598;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tetracycline resistance protein tetM from transposon Tn916
DE (TetM(916)).
GN TETM OR TET(M).
OS Enterococcus faecalis (Streptococcus faecalis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Enterococcaceae;
OC Enterococcus.
OX NCBI_TaxID=1351;
RN [1]
RP SEQUENCE FROM N.A.
RC TRANSPOSOM-Tn916;
RX MEDLINE=91045089; PubMed=2172929;
RA Burdett V.;
RT Nucleotide sequence of the tet(M) gene of Tn916.;
RL Nucleic Acids Res. 18:6137-6137(1990).
CC -|- FUNCTION: ABOLISH THE INHIBITORY EFFECT OF TETRACYCLIN ON PROTEIN
CC SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.
CC -|- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC TETM/TETO SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X56353; CAA39796.1; -.
DR PIR; S13142; S13142.
```

```
DR HSSP; P13551; IFNM.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR000795; GTP_EFTU.
DR InterPro; IPR004161; GTP_EFTU_D2.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR PROSITE; PS00301; EFACOR_GTP; 1.
KW Protein biosynthesis; Antibiotic resistance; GTP-binding;
KW Transposable element.
FT NP_BIND 10 17 GTP (BY SIMILARITY).
FT NP_BIND 74 78 GTP (BY SIMILARITY).
FT NP_BIND 128 131 GTP (BY SIMILARITY).
SQ SEQUENCE 639 AA; 72464 MW; B2315A37B53B18FB CRC64;

Query Match 63.3%; Score 31; DB 1; Length 639;
Best Local Similarity 55.6%; Pred. NO. 1e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 426 KNAEYTIHI 434
| | | | |
| | | | |

RESULT 33
TETM_NEIME STANDARD; PRT; 639 AA.
ID TETM_NEIME STANDARD; PRT; 639 AA.
AC Q51238;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Tetracycline resistance protein tetM (Tet(M)).
GN TETM OR TET(M).
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=23905;
RX MEDLINE=95247617; PubMed=7730215;
RA Gascoyne-Binzi D.M., Heritage J., Hawkey P.M., Sprott M.S.;
RT "Characterisation of a tet(M)-carrying plasmid from Neisseria
RT meningitidis.".
RL J. Antimicrob. Chemother. 34:1015-1023(1994).
CC -|- FUNCTION: ABOLISH THE INHIBITORY EFFECT OF TETRACYCLIN ON PROTEIN
CC SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.
CC -|- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC TETM/TETO SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X75073; CAA52967.1; -.
DR HSSP; P13551; IFNM.
DR InterPro; IPR000640; EFG_C.
DR InterPro; IPR000795; GTP_EFTU.
DR InterPro; IPR004161; GTP_EFTU_D2.
DR Pfam; PF00679; EFG_C; 1.
DR Pfam; PF00009; GTP_EFTU; 1.
DR Pfam; PF03144; GTP_EFTU_D2; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR PROSITE; PS00301; EFACOR_GTP; 1.
KW Protein biosynthesis; Antibiotic resistance; GTP-binding; Plasmid.
FT NP_BIND 10 17 GTP (BY SIMILARITY).
FT NP_BIND 74 78 GTP (BY SIMILARITY).
FT NP_BIND 128 131 GTP (BY SIMILARITY).
```


SQ SEQUENCE 639 AA: 72603 MW: 1D641087E280A83D CRC64;

Query Match 63.3%; Score 31; DB 1; Length 639;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
| |||:
Db 426 KNAEYTIHI 434

RESULT 34

TTM_STAAU
ID TETM_STAAU STANDARD; PRT: 639 AA.
AC Q53770;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Tetracycline resistance protein tetM (TetA(M)).
GN TETM OR TET(A(M)).
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91158314; PubMed=2073121;
RA Nesin M., Svec P., Lupski J.R., Godson G.N., Kreiswirth B.,
Projan S.J.;
RT "Cloning and nucleotide sequence of a chromosomally encoded
tetracycline resistance determinant, tetA(M), from a pathogenic,
methicillin-resistant strain of Staphylococcus aureus.";
RT Antimicrob. Agents Chemother. 34:2273-2276(1990).
RL FUNCTION: ABOLISH THE INHIBITORY EFFECT OF TETRACYCLIN ON PROTEIN
SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.
CC -!- SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC TETM/TETO SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M21136; AAA26678.1; -
CC HSSP; P13551; IFNM.
CC InterPro; IPR000640; EFG_C.
CC InterPro; IPR000795; GTP_EFTU.
CC InterPro; IPR004161; GTP_EFTU_D2.
CC Pfam; PF00679; EFG_C; 1.
CC Pfam; PF00009; GTP_EFTU; 1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC PRINTS; PR00315; ELONGATNFCT.
CC PROSITE; PS00315; ELONGATNFCT.
CC Protein biosynthesis; Antibiotic resistance; GTP-binding.
FT NP_BIND 10 17 GTP (BY SIMILARITY).
FT NP_BIND 74 78 GTP (BY SIMILARITY).
FT NP_BIND 128 131 GTP (BY SIMILARITY).
SQ SEQUENCE 639 AA: 72639 MW: 66470052A673BE1F CRC64;

Query Match 63.3%; Score 31; DB 1; Length 639;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
QY 1 KMNEYTVHL 9
| |||:
Db 426 KNAEYTIHI 434

SQ SEQUENCE 639 AA: 72603 MW: 1D641087E280A83D CRC64;

Query Match 63.3%; Score 31; DB 1; Length 639;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
| |||:
Db 426 KNAEYTIHI 434

RESULT 35

TETM_UREUR

ID TETM_UREUR STANDARD; PRT: 639 AA.
AC P09757;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Tetracycline resistance protein tetM.
GN TETM.
OS Ureaplasma urealyticum (Ureaplasma urealyticum biotype 2).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=2130;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88144009; PubMed=3344217;
RA Sanchez-Pescador R., Brown J.T., Roberts M., Urdea M.S.;
RT "The nucleotide sequence of the tetracycline resistance determinant
tetM from Ureaplasma urealyticum.";
RL Nucleic Acids Res. 16:1216-1217(1988).
CC -!- FUNCTION: ABOLISH THE INHIBITORY EFFECT OF TETRACYCLIN ON PROTEIN
SYNTHESIS BY A NON-COVALENT MODIFICATION OF THE RIBOSOMES.
CC -!- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC TETM/TETO SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U08812; AAA73978.1; -
CC PIR; S03268; S03268.
CC HSSP; P13551; IFNM.
CC InterPro; IPR000640; EFG_C.
CC InterPro; IPR000795; GTP_EFTU.
CC InterPro; IPR004161; GTP_EFTU_D2.
CC Pfam; PF00679; EFG_C; 1.
CC Pfam; PF00009; GTP_EFTU; 1.
CC Pfam; PF03144; GTP_EFTU_D2; 1.
CC PRINTS; PR00315; ELONGATNFCT.
CC PROSITE; PS00301; EFATOR_GTP; 1.
CC Protein biosynthesis; Antibiotic resistance; GTP-binding.
FT NP_BIND 10 17 GTP (BY SIMILARITY).
FT NP_BIND 74 78 GTP (BY SIMILARITY).
FT NP_BIND 128 131 GTP (BY SIMILARITY).
SQ SEQUENCE 639 AA: 72599 MW: A1497055BB182B3A CRC64;

Query Match 63.3%; Score 31; DB 1; Length 639;
Best Local Similarity 55.6%; Pred. No. 1e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
| |||:
Db 426 KNAEYTIHI 434

RESULT 36

PTK2_YEAST
ID PTK2_YEAST STANDARD; PRT: 818 AA.
AC P47116;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine/threonine-protein kinase PTK2/STK2 (EC 2.7.1.-).
GN PTK2 OR STK2 OR XJR059W OR J1725.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97079202; PubMed=8920934;

RA Nozaki T., Nishimura K., Michael A.J., Maruyama T., Kakinuma Y.,
RA Igarashi K.;
RT "A second gene encoding a putative serine/threonine protein kinase
RT which enhances spermine uptake in *Saccharomyces cerevisiae*.";
RL Biochem. Biophys. Res. Commun. 228:452-458(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=5288C;
RX MEDLINE=96437976; PubMed=8840504;
RA Huang M.-E., Manus V., Chuat J.-C., Galibert F.;
RT "Analysis of a 62 kb DNA sequence of chromosome X reveals 36 open
RT reading frames and a gene cluster with a counterpart on chromosome
RT XI.";
RL Yeast 12:869-875(1996).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=97299648; PubMed=9154797;
RA Kaouass M., Audette M., Ramotar D., Verma S., de Montigny D.,
RA Gamache I., Torossian K., Poulin R.;
RT "The STK2 gene, which encodes a putative Ser/Thr protein kinase, is
RT required for high-affinity spermidine transport in *Saccharomyces*
RT *cerevisiae*.";
RL Mol. Cell. Biol. 17:2994-3004(1997).
CC -!- FUNCTION: ESSENTIAL DETERMINANT FOR HIGH-AFFINITY SPERMIDINE
CC TRANSPORT.
CC -!- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D87274; BAAL3325.1; -;
DR EMBL; 249559; CAA89587.1; -;
DR EMBL; L47993; AAB3285.1; -;
DR HSP; P02632; ICB1.
DR SGD; S0003820; PTK2.
DR InterPro; IPR000719; Euk_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferase; Serine/threonine-protein kinase; ATP-binding.
FT DOMAIN 255 562 PROTEIN KINASE.
FT NP_BIND 261 269 ATP (BY SIMILARITY).
FT BINDING 285 285 ATP (BY SIMILARITY).
FT ACT_SITE 388 388 BY SIMILARITY.
SQ SEQUENCE 818 AA; 91400 MW; 55B6C882894FB4BD CRC64;

Query Match 63.3%; Score 31; DB 1; Length 818;
Best Local Similarity 62.5%; Pred. No. 1.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
DB 241 MKQISVHL 248

RESULT 37
COPG CAEEL
ID COPG CAEEL STANDARD; PRT; 870 AA.
AC Q22498;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable coatomer gamma subunit (Gamma-coat protein) (Gamma-COP).
GN T14G10.5.
OS *Caenorhabditis elegans*.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; *Caenorhabditis*.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wild A.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THE COATOMER IS A CYTOSOLIC PROTEIN COMPLEX THAT BINDS
CC TO DILYSINE MOTIFS AND REVERSIBLY ASSOCIATES WITH GOLGI NON-
CC CLATHRIN-COATED VESICLES, WHICH FURTHER MEDIATE BIOSYNTHETIC
CC PROTEIN TRANSPORT FROM THE ER, VIA THE GOLGI UP TO THE TRANS GOLGI
CC NETWORK. COATOMER COMPLEX IS REQUIRED FOR BUDDING FROM GOLGI
CC MEMBRANES, AND IS ESSENTIAL FOR THE RETROGRADE GOLGI-TO-ER
CC TRANSPORT OF DILYSINE-TAGGED PROTEINS (BY SIMILARITY).
CC -!- SUBUNIT: OLIGOMERIC COMPLEX THAT CONSISTS OF AT LEAST THE ALPHA,
CC BETA, BETA', GAMMA, DELTA, EPSILON AND ZETA SUBUNITS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: THE COATOMER IS CYTOPLASMIC OR POLYMERIZED
CC ON THE CYTOPLASMIC SIDE OF THE GOLGI, AS WELL AS ON THE
CC VESICLES/BUDS ORIGINATING FROM IT (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE COPG FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z68880; CAA93095.1; -;
DR WormPep; T14G10.5; CE06451.
DR InterPro; IPR002553; Adaptin_N.
DR Pfam; PF01602; Adaptin_N; 1.
KW Transport; Protein transport; Golgi stack; Membrane.
SQ SEQUENCE 870 AA; 96302 MW; 24CCA86160A60049 CRC64;

Query Match 63.3%; Score 31; DB 1; Length 870;
Best Local Similarity 83.3%; Pred. No. 1.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9
DB 638 EYTVHM 643

RESULT 38
CNA STAAU
ID CNA STAAU STANDARD; PRT; 1183 AA.
AC Q53654;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen adhesin precursor.
DE CNA.
OS *Staphylococcus aureus*.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FDA 574;
RX MEDLINE=92165839; PubMed=1311320;
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RT "Molecular characterization and expression of a gene encoding a
RT *Staphylococcus aureus* collagen adhesin".
RL J. Biol. Chem. 267:4766-4772(1992).
RN [2]
RP ERRATUM.
RA Patti J.M., Jonsson H., Guss B., Switalski L.M., Wiberg K.,
RA Lindberg M., Hoeoek M.;
RL J. Biol. Chem. 269:11672-11672(1994).

shares strong homology with transcriptional antiterminators.";
 [2]
 J. Bacteriol. 172:3966-3973(1990).
 RY SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnold M., Coudart M.P., Gonzales W.,
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
 RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
 RA Rapoport G., Danchin A.;
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees.";
 RL Mol. Microbiol. 10:371-384(1993).
 CC -1- FUNCTION: MEDIATES POSITIVE REGULATION OF THE SACPA OPERON BY
 CC FUNCTIONING AS AN ANTITERMINATOR FACTOR OF TRANSCRIPTION.
 CC -1- PTM: PHOSPHORYLATED AND INACTIVATED BY SACP (EII-SCR) (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE TRANSCRIPTIONAL ANTITERMINATOR BGLG
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J03006; AAA22726.1; ALT_INIT.
 DR EMBL: X73124; CAA51603.1; -
 DR EMBL: 299123; CAB15833.1; -
 DR HSP: F15401; IAUU.
 DR Subtilist; BG10593; sacT.
 DR InterPro: IPR001550; BglG_antitermin.
 DR Pfam: PF00874; BglG_antitermin; 2.
 DR Pfam: PF03123; CAT_RBD; 1.
 DR PROSITE: PS00654; ANTITERMINATORS_BGLG; 1.
 KW Transcription regulation; Activator; RNA-binding; Phosphorylation;
 KW Complete proteome.
 FT VARIANT 96 96 D -> Y (IN SACT30 MUTANT).
 SQ SEQUENCE 276 AA; 32074 MW; F1D63E4BC7CFBA03 CRC64;
 Query Match 61.2%; Score 30; DB 1; Length 276;
 Best Local Similarity 71.4%; Pred. No. 63;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 3 NEYTVHL 9
 Db 250 NEYQLHL 256
 |||:|
 RESULT 41
 FTR_METKA
 ID FTR_METKA STANDARD; PRT; 296 AA.
 AC Q49610;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Formylmethanofuran--tetrahydromethanopterin formyltransferase
 DE (EC 2.3.1.101).
 GN FTR.
 OS Methanopyrus kandleri.
 OC Archaea; Euryarchaeota; Methanopyrales; Methanopyrus.
 OX NCBI_TaxID=2320;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 6324;
 RX MEDLINE=95324581; PubMed=7601152;
 RA Shima S., Weiss D.S., Thauer R.K.;
 RT "Formylmethanofuran:tetrahydromethanopterin formyltransferase (Ftr)
 RT from the hyperthermophilic Methanopyrus kandleri. Cloning, sequencing
 RT and functional expression of the ftr gene and one-step purification
 RT of the enzyme overproduced in Escherichia coli.";

Eur. J. Biochem. 230:906-913(1995).
 [2]
 RY SEQUENCE OF 1-49, AND CHARACTERIZATION.
 RX MEDLINE=93130924; PubMed=1483480;
 RA Breitung J., Borner G., Scholz S., Linder D., Stetter K.O.,
 RA Thauer R.K.;
 RT "Salt dependence, kinetic properties and catalytic mechanism of N-
 RT formylmethanofuran:tetrahydromethanopterin formyltransferase from the
 RT extreme thermophile Methanopyrus kandleri.";
 RL Eur. J. Biochem. 210:971-981(1992).
 [3]
 RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS).
 RX MEDLINE=97341227; PubMed=9195883;
 RA Ermiler U., Merckel M., Thauer R., Shima S.;
 RT "Formylmethanofuran:tetrahydromethanopterin formyltransferase from
 RT Methanopyrus kandleri - new insights into salt-dependence and
 RT thermostability.";
 RL Structure 5:635-646(1997).
 CC -1- CATALYTIC ACTIVITY: N-formylmethanofuran + 5,6,7,8-
 CC tetrahydromethanopterin = methanofuran + 5-formyl-5,6,7,8-
 CC tetrahydromethanopterin.
 CC -1- PATHWAY: INVOLVED IN THE FORMATION OF METHANE FROM CO(2).
 CC -1- SUBUNIT: MONOMER.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X85115; CAA59435.1; -
 DR PUB: 1FTR; 25-NOV-98.
 DR InterPro: IPR002770; FTR.
 DR Pfam: PF01913; FTR; 1.
 DR Pfam: PF02741; FTR_C; 1.
 DR ProDom: PD007702; FTR; 1.
 KW Transferase; Acyltransferase; Methanogenesis; 3D-structure.
 FT CONFLICT 30 30 H -> D (IN REF. 2).
 FT CONFLICT 32 32 W -> K (IN REF. 2).
 FT CONFLICT 39 39 E -> K (IN REF. 2).
 SQ SEQUENCE 296 AA; 31661 MW; DDE02D3E7D98FC86 CRC64;
 Query Match 61.2%; Score 30; DB 1; Length 296;
 Best Local Similarity 44.4%; Pred. No. 68;
 Matches 4; Conservative 3; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KMNEYTVHL 9
 Db 284 KLGQYEIHL 292
 |:|:|
 RESULT 42
 K6P1_VIBCH STANDARD; PRT; 320 AA.
 ID K6P1_VIBCH
 AC Q9KNP2;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE 6-phosphofructokinase isozyme I (EC 2.7.1.11) (Phosphofructokinase-1)
 DE (Phosphohexokinase-1).
 GN PFKA OR VC2689.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
 RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

```
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
CC -1- CATALYTIC ACTIVITY: ATP + D-fructose 6-phosphate = ADP + D-
CC fructose 1,6-bisphosphate.
CC -1- ENZYME REGULATION: ACTIVATED BY ADP AND INHIBITED BY
CC PHOSPHOENOLPYRUVATE (BY SIMILARITY).
CC -1- PATHWAY: KEY CONTROL STEP OF GLYCOLYSIS.
CC -1- SUBUNIT: HOMOTETRAMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOFRUCTOKINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE004334; AAF95830.1; ALT_INIT.
DR TIGR; VC2689; .
DR InterPro; IPR000023; Phosphofructokinase.
DR Pfam; PF00365; PFK; 1.
DR PRINTS; PR00476; PHFRCTKINASE.
DR PRODOM; PD000707; Phosphofructokinase; 1.
DR PROSITE; PS00433; PHOSPHOFRUCTOKINASE; 1.
KW Kinase; Transferase; Glycolysis; Complete proteome.
SQ SEQUENCE 320 AA; 34670 MW; 552D3565CB63EB76 CRC64;

Query Match 61.2%; Score 30; DB 1; Length 320;
Best Local Similarity 55.6%; Pred. No. 75;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 267 RMGNVAVHL 275
:| | | | |

RESULT 43
PE10_CAEEL STANDARD; PRT; 407 AA.
AC P41991;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protein pes-10.
GN PES-10 OR Y46G5A.27.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Seydoux G.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Wallis J.M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U15304; AAS0369.1; .
DR EMBL; AL110485; CAB60370.1; .
DR WormPep; Y46G5A.27; CE21976.
SQ SEQUENCE 407 AA; 47017 MW; 2E0712B846FF010 CRC64;

Query Match 61.2%; Score 30; DB 1; Length 407;
Best Local Similarity 71.4%; Pred. No. 97;
Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NEYTVHL 9
DB 4 NEYMMHL 10
:| | | | |

RESULT 44
NSMA_CAEEL STANDARD; PRT; 445 AA.
AC Q45870;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Putative neutral sphingomyelinase (EC 3.1.4.12).
GN T27F6.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Dobson R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: Sphingomyelin + H(2)O = N-acylsphingosine +
CC choline phosphate.
CC -1- SIMILARITY: BELONGS TO THE NEUTRAL SPHINGOMYELINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z82060; CAB04885.1; .
DR WormPep; T27F6.6; CE16515.
KW Hypothetical protein; Hydrolase; Transmembrane.
FT TRANSMEM 377 399 POTENTIAL.
FT TRANSMEM 403 425 POTENTIAL.
FT METAL 83 83 MAGNESIUM (BY SIMILARITY).
FT SITE 215 215 IMPORTANT FOR SUBSTRATE RECOGNITION (BY
FT ACT_SITE 318 318 SIMILARITY).
FT SEQUENCE 445 AA; 51054 MW; 969FEC23A037FE72 CRC64;

Query Match 61.2%; Score 30; DB 1; Length 445;
Best Local Similarity 55.6%; Pred. No. 11e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 164 RVNFTYTHL 172
:| | | | |

RESULT 45
MURF_BORBU STANDARD; PRT; 464 AA.
ID MURF_BORBU
AC Q44777; O51284;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
```

```

DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylmuramoyl-D-glutamyl-2,6-diaminopimelate--D-alanyl-D-
DE alanyl ligase (EC 6.3.2.15) (UDP-MurNAC-pentapeptide synthetase)
DE (D-alanyl-D-alanine-adding enzyme).
GN MURF OR BB0304.
OS Borrelia burgdorferi (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; Borrelia.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE=98065943; PubMed=9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA van Vugt R., Palmer N., Adams M.D., Cocayne J.D., Weidman J.,
RA Utterback T., Wathey L., McDonald L., Artiach P., Bowman C.,
RA Smith H.O., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA "Genomic sequence of a Lyme disease spirochaete, Borrelia
RT burgdorferi."
RL Nature 390:580-586(1997).
RN [2]
RP SEQUENCE OF 67-464 FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RA Dunn J.J., Butler-Loffredo L., Kieleczawa J., Medalle J., Luft B.J.;
RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN CELL WALL FORMATION. CATALYSES THE FINAL
CC STEP IN THE SYNTHESIS OF UDP-N-ACETYLMURAMOYL-PENTAPEPTIDE, THE
CC PRECURSOR OF MUREIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + UDP-N-acetylmuramoyl-L-alanyl-D-
CC glutamyl-meso-2,6-diaminoheptanedioate + D-alanyl-D-alanine = ADP
CC + phosphate + UDP-N-acetylmuramoyl-L-alanyl-D-glutamyl-6-
CC carboxy-L-lysyl-D-alanyl-D-alanine.
CC -!- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: BELONGS TO THE MURCDEF FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001137; AAC56644.1; -.
DR EMBL; U43739; AAA85627.1; -.
DR TIGR; BB0304; -.
DR InterPro; IPR000713; Mur_ligase.
DR InterPro; IPR004101; Mur_ligase_C.
DR Pfam; PF01225; Mur_ligase_1.
DR Pfam; PF02875; Mur_ligase_C_1.
DR Peptidoglycan synthesis; Cell division; Cell wall; Ligase;
KW ATP-binding; Complete proteome.
FT NP_BIND 125 131 ATP (POTENTIAL).
FT CONFLICT 462 464 IFR -> YI (IN REF. 2).
SQ SEQUENCE 464 AA; 53446 MW; 2903850C533799BC CRC64;

Query Match 61.2%; Score 30; DB 1; Length 464;
Best Local Similarity 55.6%; Pred. NO. 1.1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db :||:|:|
235 EMNDYCVYL 243

```

Search completed: November 6, 2002, 12:09:01
Job time : 8.3333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 23.7778 Seconds
(without alignments)
42.042 Million cell updates/sec

Title: US-09-905-083-31
Perfect score: 49
Sequence: 1 KMNEYTVHL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	49	100.0	9	AAE08236	Human stratum corn
2	49	100.0	9	AAE08294	Human stratum corn
3	49	100.0	136	ABG23378	Novel human diagno
4	49	100.0	225	AAAB98502	Human stratum corn
5	49	100.0	253	AAAR67888	Human stratum corn
6	49	100.0	253	AAAW05383	Human amyloid prec
7	49	100.0	257	AAAB21326	Human HSCCE.
8	39	79.6	9	AAE08326	Human stratum corn
9	37	75.5	9	AAE08298	Human stratum corn
10	37	75.5	334	AAU37656	Streptococcus pneu
11	35	71.4	434	AAAM0085	Human protein SEQ

12	35	71.4	685	22	ABG18060	Novel human diagno
13	35	71.4	685	22	AAAM79101	Human protein SEQ
14	35	71.4	744	22	ABG18061	Novel human diagno
15	33	67.3	358	22	AAG82521	S. epidermidis ope
16	33	67.3	365	22	AAG81688	S. epidermidis ope
17	33	67.3	799	21	AAAB08074	Amino acid sequenc
18	33	67.3	804	21	AAAB08073	Plasmodium falcipa
19	33	67.3	1308	21	AAAB18167	Human novel foetal
20	32	65.3	63	22	AAU20718	Phaffia derived gl
21	32	65.3	121	18	AAW22495	Arabidopsis thalia
22	32	65.3	276	21	AAAG14213	Arabidopsis thalia
23	32	65.3	284	21	AAAG47496	Arabidopsis thalia
24	32	65.3	316	21	AAAG14212	Arabidopsis thalia
25	32	65.3	324	21	AAAG47495	Arabidopsis thalia
26	32	65.3	364	21	AAAG14211	Arabidopsis thalia
27	32	65.3	372	21	AAAG47494	Arabidopsis thalia
28	32	65.3	398	22	AAU33973	Staphylococcus aur
29	32	65.3	409	22	AAU36825	Tetracycline resis
30	32	65.3	639	9	AAAP80147	Drosophila melanoq
31	32	65.3	678	22	ABB65057	Novel human diagno
32	32	65.3	683	22	ABG06295	Human polypeptide
33	32	65.3	688	22	AAAB39237	Novel protein kina
34	32	65.3	688	22	AAAB65597	Novel human diagno
35	32	65.3	692	22	ABG08784	Human polypeptide
36	32	65.3	698	22	AAAM41023	Human polypeptide
37	32	65.3	752	22	AAU37828	Streptococcus pneu
38	32	65.3	752	22	AAU38068	Streptococcus pneu
39	32	65.3	796	22	ABG01448	Novel human diagno
40	32	65.3	796	22	ABG11797	Novel human diagno
41	32	65.3	882	22	ABG05821	Novel human diagno
42	32	65.3	917	22	ABG06296	Novel human diagno
43	32	65.3	917	22	ABG08780	Novel human diagno
44	32	65.3	951	22	AAAM79950	Human protein SEQ
45	32	65.3	1015	22	ABG08785	Novel human diagno

ALIGNMENTS

RESULT 1
AAE08236
ID AAE08236 standard; peptide; 9 AA.
XX
AC AAE08236;
XX
DT 01-NOV-2001 (first entry)
XX
XX Human stratum corneum chymotrypsin enzyme peptide #1 (residues 72-80).
DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.
XX
OS Homo sapiens.
XX
PN WO200159158-A1.
XX
PD 16-AUG-2001.
XX
PF 07-FEB-2001; 2001WO-US03977.
XX
PR 11-FEB-2000; 2000US-0502600.
XX
PA (UYAR-) UNIV ARKANSAS.
XX
PI O'brien TJ;
XX
DR WPI; 2001-514676/56.
XX
PT Diagnosing cancer comprises detecting stratum corneum chymotrypsin
enzyme -
XX
PS Claim 25; Page 102; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.

XX Sequence 9 AA;

Query Match 100.0%; Score 49; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
 DB 1 KMNEYTVHL 9
 |||||

RESULT 2

AAE08294
 ID AAE08294 standard; peptide; 9 AA.

AC AAE08294;

DT 01-NOV-2001 (first entry)

XX Human stratum corneum chymotrypsin enzyme peptide #59 (residues 72-80).
 XX Stratum corneum chymotrypsin enzyme; SCCE: cytostatic; vaccine; tumour;
 KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
 KW antisense therapy; malignant hyperplasia.

XX Homo sapiens.

XX WO200159158-A1.

PD 16-AUG-2001.

XX 07-FEB-2001; 2001WO-US03977.

PR 11-FEB-2000; 2000US-0502600.

XX (UYAR-) UNIV ARKANSAS.

XX O'brien TJ;

XX WPI; 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
 PT enzyme -

PS Disclosure; Page 115; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.

XX Sequence 9 AA;

Query Match 100.0%; Score 49; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
 DB 1 KMNEYTVHL 9
 |||||

RESULT 3

ABG23378
 ID ABG23378 standard; Protein; 136 AA.

XX AC ABG23378;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #23369.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR N-PSDB; AAS87565.

XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX Claim 20; SEQ ID NO 53737; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 136 AA;

Query Match 100.0%; Score 49; DB 22; Length 136;
 Best Local Similarity 100.0%; Pred. No. 0.059;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
 |||||

Db 78 KMNEYTVHL 86

RESULT 4

AAB98502

ID AAB98502 standard; Protein; 225 AA.

XX

AC AAB98502;

XX

DT 03-AUG-2001 (first entry)

XX

DE Human Stratum Corneum Chymotryptic Enzyme, SCCE, catalytic domain.

XX

KW Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;

KW tumour antigen-derived gene 15; serine protease;

KW

KW Stratum Corneum Chymotryptic Enzyme; SCCE.

XX

OS Homo sapiens.

XX

PN WO200129056-A1.

XX

PD 26-APR-2001.

XX

PF 20-OCT-2000; 2000WO-US29095.

XX

PR 20-OCT-1999; 98US-0421213.

XX

PA (UYAR-) UNIV ARKANSAS.

XX

PI O'Brien TJ, Tanimoto H;

XX

DR WPI; 2001-381031/40.

XX

PT Novel extracellular serine protease, termed tumor antigen-derived gene

PT

PT 15 protein overexpressed in carcinomas and DNA encoding it, for

PT

PT diagnosis, treatment, prevention of cancer, particularly breast,

PT

PS ovarian cancer -

XX

PS Example 10; Fig 1; 130pp; English.

XX

CC The present invention relates to human tumour antigen-derived gene 15

CC

CC (TADG-15) protein and coding sequence (see AAH23601 and AAB98500).

CC

CC TADG-15 is an extracellular serine protease. It was found that TADG-15 is

CC

CC over-expressed in ovarian tumours. TADG-15 protein or its fragments of

CC

CC 9-20 residues that lack TADG-15 protease activity are useful for

CC

CC vaccinating an individual against TADG-15, having, suspected of having or

CC

CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a

CC

CC diagnostic or therapeutic target in cancer. The present sequence was used

CC

CC in a sequence homology alignment with the catalytic domain of TADG-15.

XX

SQ Sequence 225 AA;

XX

Query Match

Best Local Similarity

Matches

9; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY

1 KMNEYTVHL 9

IIIIIIIIII

Db

44 KMNEYTVHL 52

IIIIIIIIII

RESULT 5

AAR67888

ID AAR67888 standard; Protein; 253 AA.

XX

AC AAR67888;

XX

DT 09-AUG-1995 (first entry)

XX

DE Human stratum corneum chymotrophic recombinant enzyme (SCCE).

XX

KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;

KW

callosities; keratosis pilaris; ichthyoses; eczema.

XX Homo sapiens.

XX OS

XX PN WO9500651-A.

XX XX

XX PD 05-JAN-1995.

XX XX

XX PF 20-JUN-1994; 94WO-IB00166.

XX XX

XX PR 18-JUN-1993; 93DK-0000725.

XX XX

XX PA (SYMB-) SYMBICOM AB.

XX XX

XX PI Egelrud T, Hansson L;

XX XX

XX DR WPI; 1995-052088/07.

XX XX

XX DR N-PSDB; AAQ81203.

XX XX

XX PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme

XX

XX - and related vectors, transformed cells and polypeptides, and

XX

XX useful for treating skin disorders, e.g. acne or psoriasis, and

XX

XX for identification of specific inhibitors.

XX

XX PS Disclosure; Page 97; 137pp; English.

XX XX

XX CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic

XX

XX and skin care products, especially to treat and prevent acne,

XX

XX xeroderma, or other hyperkeratotic conditions (e.g. callosities or

XX

XX keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is

XX

XX CC produced recombinantly following mammal, insect, plant, or

XX

XX CC microorganism transformation with plasmid pS507.

XX

XX SQ Sequence 253 AA;

XX

XX Query Match

Best Local Similarity

Matches

9; Conservative

0; Mismatches

0; Indels

0; Gaps

0;

QY

1 KMNEYTVHL 9

IIIIIIIIII

Db

72 KMNEYTVHL 80

IIIIIIIIII

RESULT 6

AAR05383

ID AAR05383 standard; Protein; 253 AA.

XX

XX AC AAR05383;

XX XX

XX DT 31-DEC-1996 (first entry)

XX XX

XX DE Human amyloid precursor protein protease.

XX XX

XX KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;

XX

XX KW therapy.

XX OS

XX Homo sapiens.

XX XX

XX PN WO9631122-A1.

XX XX

XX PD 10-OCT-1996.

XX XX

XX PF 02-APR-1996; 96WO-US04294.

XX XX

XX PR 04-APR-1995; 95US-0416257.

XX XX

XX PA (ELIL) LILLY & CO ELI.

XX

XX PI Dixon EP, Johnstone EM, Little SP;

XX

XX WPI; 1996-464694/46.

XX

XX DR N-PSDB; AAT39783.

XX

XX XX

PT New isolated human amyloid precursor protein protease - used to
 PT develop prods. for the treatment or diagnosis of associated
 PT conditions, esp. Alzheimer's disease

XX Claim 1; Page 44-45; 55pp; English.

XX Human amyloid precursor protein protease (AAW05383) is involved in
 CC the processing or clearance of amyloid precursor protein to form
 CC beta-amyloid peptide. Its amino acid sequence was deduced from
 CC a cDNA clone (AAT39783) obtd. from a human lung library. Recombinant
 CC protease can be produced in transformed or transfected prokaryotic
 CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
 CC used to develop products for the design and testing of cpds. useful
 CC for treating or preventing conditions associated with beta-amyloid
 CC peptide, esp. Alzheimer's disease.

XX Sequence 253 AA;

Query Match 100.0%; Score 49; DB 17; Length 253;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

Db 72 KMNEYTVHL 80

RESULT 7

AAB21326
 ID AAB21326 standard; Protein; 257 AA.

XX AC AAB21326;

DT 02-FEB-2001 (first entry)

XX Human HSCEE.

XX Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCEE;
 KW Human stratum corneum chymotryptic enzyme; kallikrein-like protein;
 KW serine protease; cytostatic; cancer; prostate cancer.

XX Homo sapiens.

XX WO200053776-A2.

XX 14-SEP-2000.

PF 09-MAR-2000; 2000WO-CA00258.

XX 11-MAR-1999; 99US-0124260.

PR 01-APR-1999; 99US-0127386.

PR 21-JUL-1999; 99US-0144919.

PA (MOUN) MOUNT SINAI HOSPITAL.

XX Yousef GM, Diamandis EP;

XX WPI; 2000-587440/55.

XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
 PT protein mediated disorders, especially cancer.

PS Example 4; Fig 17; 184pp; English.

XX The present sequence is human stratum corneum chymotryptic enzyme
 CC (HSCEE), a member of the kallikrein multi-gene family. Kallikreins and
 CC kallikrein-like proteins are a subgroup of the serine protease enzyme
 CC family. They catalyse the selective cleavage of specific polypeptide
 CC precursors to release peptides with potent biological activity. Nucleic
 CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
 CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
 CC treatment, monitoring and diagnosis of cancers, especially prostate
 CC cancer. They can also be used to identify a substance that can associate

CC with or mediate the biological activity of the proteins. Antibodies can
 CC be used to treat conditions mediated by the kallikrein-like proteins.

SQ Sequence 257 AA;

Query Match 100.0%; Score 49; DB 21; Length 257;
 Best Local Similarity 100.0%; Pred. No. 0.12;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

Db 77 KMNEYTVHL 85

RESULT 8

AAE08326
 ID AAE08326 standard; peptide; 9 AA.

XX AC AAE08326;

DT 01-NOV-2001 (first entry)

XX Human stratum corneum chymotrypsin enzyme peptide #91 (residues 74-82).

XX Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
 KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
 KW antisense therapy; malignant hyperplasia.

XX Homo sapiens.

XX WO200159158-A1.

XX 16-AUG-2001.

PF 07-FEB-2001; 2001WO-US03977.

XX 11-FEB-2000; 2000US-0502600.

XX (UYAR-) UNIV ARKANSAS.

XX O'brien TJ;

XX WPI; 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
 PT enzyme

XX Disclosure; Page 122; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.

SQ Sequence 9 AA;

Query Match 79.6%; Score 39; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 NEYTVHL 9

Db 1 NEYTVHL 7

RESULT 9

AAE08298

ID AAE08298 standard; peptide; 9 AA.

```

XX AAE08298;
AC
XX
XX 01-NOV-2001 (first entry)
DT
XX
XX Human stratum corneum chymotrypsin enzyme peptide #63 (residues 70-78).
DE
XX Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.
XX
XX Homo sapiens.
OS
XX WO200159158-A1.
XX
XX 16-AUG-2001.
PD
XX
XX 07-FEB-2001; 2001WO-US03977.
PF
XX
XX 11-FEB-2000; 2000US-0502600.
PR
XX
XX (UYAR-) UNIV ARKANSAS.
PA
XX O'brien TJ;
PI
XX
XX WPI; 2001-514576/56.
DR
XX
XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -
PT
XX
XX Disclosure; Page 116; 127pp; English.
PS
XX
XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.
XX
XX Sequence 9 AA;
SQ
Query Match 75.5%; Score 37; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTV 7
DB 3 KMNEYTV 9
|||||
|||||

RESULT 10
AAU37656
ID AAU37656 standard; Protein; 334 AA.
AC
XX
XX AAU37656;
XX
XX 14-FEB-2002 (first entry)
DT
XX
XX Streptococcus pneumoniae cellular proliferation protein #85.
DE
XX
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
KW
XX
XX Streptococcus pneumoniae.
OS
XX WO200170955-A2.
XX
XX 27-SEP-2001.
PD
XX
XX 21-MAR-2001; 2001WO-US09180.
PF

```

```

XX 21-MAR-2000; 2000US-191078P.
PR
XX 23-MAY-2000; 2000US-206848P.
PR
XX 26-MAY-2000; 2000US-207727P.
PR
XX 23-OCT-2000; 2000US-242578P.
PR
XX 27-NOV-2000; 2000US-253625P.
PR
XX 22-DEC-2000; 2000US-257931P.
PR
XX 16-FEB-2001; 2001US-269308P.
XX
XX (ELIT-) ELITRA PHARM INC.
PA
XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI; 2001-611495/70.
DR
XX N-PSDB; AAS55515.
DR
XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
PT
XX
XX Example 3; Seq ID No 13249; 511pp; English.
PS
XX
XX The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 334 AA;
SQ
Query Match 75.5%; Score 37; DB 22; Length 334;
Best Local Similarity 66.7%; Pred. No. 34;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 302 KMNEYITHI 310
|||||
|||||

RESULT 11
AAM80085
ID AAM80085 standard; Protein; 434 AA.
XX
XX
XX AC AAM80085;
XX
XX 06-NOV-2001 (first entry)
DT
XX
XX Human protein SEQ ID NO 3731.
DE
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
XX Homo sapiens.
OS
XX WO200157190-A2.
XX
XX 09-AUG-2001.
PF

```

XX PF 05-FEB-2001; 2001WO-US04098.
XX PR 03-FEB-2000; 2000US-0496914.
XX PR 27-APR-2000; 2000US-0560875.
XX PR 20-JUN-2000; 2000US-0598075.
XX PR 19-JUL-2000; 2000US-0620325.
XX PR 01-SEP-2000; 2000US-0654936.
XX PR 15-SEP-2000; 2000US-0663561.
XX PR 20-OCT-2000; 2000US-0693325.
XX PR 30-NOV-2000; 2000US-0728422.
XX PA (HYSE-) HYSEQ INC.
XX PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
XX PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
XX N-PSDB; AAK53218.
XX Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
XX Claim 20; Page 419; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
XX encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
XX cytokine, cell proliferation or cell differentiation or which may induce
XX production of other cytokines in other cell populations. The
XX polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX peptide therapy. The polypeptides have various cytokine-like activities,
XX e.g. stem cell growth factor activity, haematopoiesis regulating
XX activity, tissue growth factor activity, immunomodulatory activity and
XX activin/inhibin activity and may be useful in the diagnosis and/or
XX treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX inflammation.
XX Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
XX (AAM80020) are omitted as the relevant pages from the sequence listing
XX were missing at the time of publication.
XX Sequence 434 AA;
XX Query Match 71.4%; Score 35; DB 22; Length 434;
XX Best Local Similarity 75.0%; Pred. No. 1.1e+02;
XX Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX QY 1 KMNEYTVH 8
XX | | | | |
XX Db 192 KQHNYTVH 199
XX
XX RESULT 12
XX AAG18060
XX ID AAG18060 standard; Protein; 685 AA.
XX AC AAG18060;
XX DT 18-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #18051.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX tissue growth factor; immunomodulatory; cancer; leukaemia;
XX nervous system disorder; arthritis; inflammation.
XX OS Homo sapiens.
XX XX WO200175067-A2.
XX PN 11-OCT-2001.
XX PD 30-MAR-2001; 2001WO-US06631.
XX PF
XX XX

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX PA Drmanac RT, Liu C, Tang YT;
XX PI WPI; 2001-639362/73.
XX DR N-PSDB; AAS82247.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity -
XX PS Claim 20; SEQ ID No 48419; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. AAG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 685 AA;
XX Query Match 71.4%; Score 35; DB 22; Length 685;
XX Best Local Similarity 75.0%; Pred. No. 1.9e+02;
XX Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX QY 1 KMNEYTVH 8
XX | | | | |
XX Db 444 KQHNYTVH 451
XX
XX RESULT 13
XX AAM79101
XX ID AAM79101 standard; Protein; 685 AA.
XX AC AAM79101;
XX DT 06-NOV-2001 (first entry)
XX DE Human protein SEQ ID NO 1763.
XX XX
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;
XX KW nervous system disorder; arthritis; inflammation.
XX OS Homo sapiens.
XX XX WO200157190-A2.
XX PN 09-AUG-2001.
XX PD 05-FEB-2001; 2001WO-US04098.
XX PF
XX XX

PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
XX WPI: 2001-476283/51.
DR N-PSDB; AAK52234.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
PT
XX
XX Claim 20; Page 4108-4109; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX
XX Sequence 685 AA;
Query Match 71.4%; Score 35; DB 22; Length 685;
Best Local Similarity 75.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
| | | | |
DB 444 KQHNYTVH 451
RESULT 14
ABG18061
ID ABG18061 standard; Protein; 744 AA.
XX
XX AC ABG18061;
XX
XX 18-FEB-2002 (first entry)
XX
XX DE Novel human diagnostic protein #18052.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX OS Homo sapiens.
XX
XX PN WO200175067-A2.
XX
XX PD 11-OCT-2001.
XX
XX PF 30-MAR-2001; 2001WO-US08631.
XX
XX PR 31-MAR-2000; 2000US-0540217.
XX 23-AUG-2000; 2000US-0649167.
XX

PA (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS82248.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 48420; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (III) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 744 AA;
Query Match 71.4%; Score 35; DB 22; Length 744;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
| | | | |
DB 503 KQHNYTVH 510
RESULT 15
AAG82521
ID AAG82521 standard; Protein; 358 AA.
XX
XX AC AAG82521;
XX
XX DT 03-SEP-2001 (first entry)
XX
XX DE S. epidermidis open reading frame protein sequence SEQ ID NO:2136.
XX
XX KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
XX OS Staphylococcus epidermidis.
XX
XX PN WO200134809-A2.
XX
XX PD 17-MAY-2001.
XX
XX PF 09-NOV-2000; 2000WO-US30782.
XX
XX PR 09-NOV-1999; 99US-0164258.
XX (GLAX) GLAXO GROUP LTD.
XX
XX PI Kimmerly WJ;

```

XX WPI; 2001-316495/33.
DR N-PSDB; AAH53371.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 576; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 358 AA;
SQ
Query Match 67.3%; Score 33; DB 22; Length 358;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 NEYTVHL 9
Db 59 NEYLVHL 65

RESULT 16
AAG81688
ID AAG81688 standard; Protein; 365 AA.
XX
AC AAG81688;
XX
DT 03-SEP-2001 (first entry)
XX
DE S. epidermidis open reading frame protein sequence SEQ ID NO:470.
XX
KW Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
OS Staphylococcus epidermidis.
XX
PN WO200134809-A2.
XX
PD 17-MAY-2001.
XX
PF 09-NOV-2000; 2000WO-US30782.
XX
PR 09-NOV-1999; 99US-0164258.
XX
PA (GLAXO) GLAXO GROUP LTD.
XX
PI Kimmerly WJ;
XX
XX WPI; 2001-316495/33.
DR N-PSDB; AAH52538.
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
XX
XX Claim 18; Page 165; 2188pp; English.
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 365 AA;
SQ
Query Match 67.3%; Score 33; DB 22; Length 365;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 3 NEYTVHL 9
Db 66 NEYLVHL 72

RESULT 17
AAB08074
ID AAB08074 standard; Protein; 799 AA.
XX
AC AAB08074;
XX
DT 04-DEC-2000 (first entry)
XX
DE Amino acid sequence of a murine GAP12 polypeptide.
XX
KW GAP12; gene therapy; cancer; cell identification.
XX
OS Mus musculus.
XX
PN US6100058-A.
XX
PD 08-AUG-2000.
XX
PF 12-AUG-1997; 97US-0909954.
XX
PR 12-AUG-1997; 97US-0909954.
XX
PA (AXYS-) AXYS PHARM INC.
XX
PI Allen MJ, Buckler AJ;
XX
DR WPI; 2000-542423/49.
DR N-PSDB; AAA63667.
XX
XX New GAP12 genes and proteins for identifying homologous genes, gene
PT therapy to treat disorders associated with GAP12 defects, mapping
PT functional regions of the protein, or in studying associated
PT physiological pathways -
XX
XX Claim 1; Column 31-36; 40pp; English.
XX
XX The present sequence represents a mammalian GAP12 protein. The GAP12
CC genes and proteins are useful in identifying homologous or related

```

CC genes, in producing compositions that modulate the expression or
 CC function of its encoded protein, for gene therapy to treat disorders
 CC associated with GAP12 defects, for mapping functional regions of the
 CC protein, and in studying associated physiological pathways. The DNA
 CC may also be used to identify expression of the gene in a biological
 CC specimen and to generate transgenic animals or site-specific gene
 CC modifications in cell lines. The polypeptides are useful for the
 CC production of antibodies. Modulation of gene activity in vivo is used
 CC for prophylactic and therapeutic purposes, such as treatment of cancer,
 CC or identification of cell type based on expression.

XX SQ Sequence 799 AA;

Query Match 67.3%; Score 33; DB 21; Length 799;
 Best Local Similarity 100.0%; Pred. No. 5.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9
 Db 56 EYTVHL 61
 |||||

RESULT 18

AAB08073
 ID AAB08073 standard; Protein; 804 AA.

XX AC AAB08073;

XX DT 04-DEC-2000 (first entry)

XX DE Amino acid sequence of a human GAP12 polypeptide.

XX KW GAP12; gene therapy; cancer; cell identification.

XX OS Homo sapiens;

XX PN US6100058-A.

XX PD 08-AUG-2000.

XX PF 12-AUG-1997; 97US-0909954.

XX PR 12-AUG-1997; 97US-0909954.

XX PA (AXYS-) AXYS PHARM INC.

XX PI Allen MJ, Buckler AJ;

XX WPI; 2000-542423/49.

XX DR N-PSDB; AAA63666.

XX PT New GAP12 genes and proteins for identifying homologous genes, gene
 PT therapy to treat disorders associated with GAP12 defects, mapping
 PT functional regions of the protein, or in studying associated
 PT physiological pathways -

XX PS Claim 1; Column 23-28; 40pp; English.

XX CC The present sequence represents mammalian GAP12 protein. The GAP12
 CC genes and proteins are useful in identifying homologous or related
 CC genes, in producing compositions that modulate the expression or
 CC function of its encoded protein, for gene therapy to treat disorders
 CC associated with GAP12 defects, for mapping functional regions of the
 CC protein, and in studying associated physiological pathways. The DNA
 CC may also be used to identify expression of the gene in a biological
 CC specimen and to generate transgenic animals or site-specific gene
 CC modifications in cell lines. The polypeptides are useful for the
 CC production of antibodies. Modulation of gene activity in vivo is used
 CC for prophylactic and therapeutic purposes, such as treatment of cancer,
 CC or identification of cell type based on expression.

XX SQ Sequence 804 AA;

Query Match 67.3%; Score 33; DB 21; Length 1308;
 Best Local Similarity 66.7%; Pred. No. 9.5e+02;

Query Match 67.3%; Score 33; DB 21; Length 804;
 Best Local Similarity 100.0%; Pred. No. 5.5e+02;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9
 Db 56 EYTVHL 61
 |||||

RESULT 19

AAB18167
 ID AAB18167 standard; Protein; 1308 AA.

XX AC AAB18167;

XX DT 07-NOV-2000 (first entry)

XX DE Plasmodium falciparum chromosome 2 related protein SEQ ID NO:24.

XX KW Plasmodium falciparum; chromosome 2; human malaria parasite; vaccine;
 KW antimalarial; malaria; protozoacide; infection; insecticide.

XX OS Plasmodium falciparum.

XX PN WO200025728-A2.

XX PD 11-MAY-2000.

XX PF 05-NOV-1999; 99WO-US26796.

XX PR 05-NOV-1998; 98US-0107131.

XX PA (HOFF/) HOFFMAN S.

XX PA (CARU/) CARUCCI D.

XX PA (GARD/) GARDNER M.

XX PA (VENT/) VENTER J C.

XX PI Hoffman S, Carucci D, Gardner M, Venter JC;

XX WPI; 2000-365347/31.

XX PT Proteins encoded by chromosome 2 of the human malarial parasite,
 PT Plasmodium falciparum, useful as antimalarial vaccines and in the
 PT diagnosis of P.falciparum infection -

XX PS Disclosure; Page 64-68; 577pp; English.

XX CC The present invention describes proteins and their fragments (I) encoded
 CC by chromosome 2 of the human malarial parasite, Plasmodium falciparum.
 CC Also described are: (1) nucleotide sequences (II) encoding (I); and (2)
 CC vaccines against P. falciparum infection comprising (I) or (II).
 CC (1) and (II) are useful for the development of vaccines against
 CC P. falciparum infection. (I) and polyclonal antisera or a monoclonal
 CC antibody raised to immunogens comprising the sequences of (I), are
 CC useful in the detection of infection with P. falciparum. Furthermore,
 CC (I) (especially when they are rifins or secreted or membrane proteins)
 CC can aid the identification of drugs to treat or prevent P. falciparum
 CC infection, or they can be used to identify drug resistance in
 CC P. falciparum. Sequencing of the Plasmodium chromosome 2 and the
 CC subsequent identification of proteins encoded by it will help to expand
 CC our understanding of parasite biology, a process hampered by the
 CC complexity of the parasitic lifecycle, and provide new targets for
 CC vaccine and drug development. Parasite resistance to drugs and mosquito
 CC resistance to insecticides have led to a resurgence of malaria in many
 CC parts of the world, and there is a pressing need for vaccines and new
 CC drugs. AAA70078 to AAA70287 and AAB18144 to AAB18352 represent nucleotide
 CC and protein sequences given in the present invention, but which are not
 CC specifically mentioned within the specification.

XX SQ Sequence 1308 AA;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 KMNEYTVHL 9
I I I I I I I;
Db 1146 KNEYTKHV 1154

RESULT 20
AAU20718
ID AAU20718 standard; Protein; 63 AA.
XX AAU20718;
XX
DT 17-DEC-2001 (first entry)
XX
DE Human novel foetal antigen, SEQ ID NO 962.
XX
KW Human; foetal tissue antigen; antinflammatory; neuroprotective;
KW immunomodulator; cardiovascular; cytostatic; nephrothropic;
KW cardiovascular; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; breast neoplasm; cancer;
KW cardiovascular disorder; cardiac arrest; cerebrovascular disorder;
KW cerebral ischaemia; angiogenesis; nervous system disorder;
KW Alzheimer's disease; infection; ocular disorder; corneal infection;
KW wound healing; epithelial cell proliferation; food additive.
XX
OS Homo sapiens.
XX
XX
PN WO20015312-A2.
XX
PD 02-AUG-2001.
XX
XX
PF 17-JAN-2001; 2001WO-US01321.
XX
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180626.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.

PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
XX WPI; 2001-488782/53.
DR N-PSDB; AAS33538.
XX
XX New polynucleotides and polypeptides for diagnosing, treating,
PT preventing or prognosing e.g. diseases or disorders of the nervous,
PT musculoskeletal, excretory, gastrointestinal, reproductive, and
PT respiratory systems -
XX
XX
XX Claim 11; SEQ ID No 962; 642pp; English.

PS
XX

CC The invention relates to novel nucleic acids encoding novel human foetal
CC antigens. The nucleic acids and proteins are used to prevent, treat (e.g.
CC by gene therapy) or ameliorate a medical condition in e.g. humans, mice,
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
CC are also used in diagnosing a pathological condition or susceptibility
CC to a pathological condition. The antibodies to the antigens can also
CC be used in alleviating symptoms associated with the disorders and in
CC diagnostic immunosays e.g. radioimmunosays or enzyme linked
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated
CC include autoimmune diseases e.g. rheumatoid arthritis,
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
CC and ocular disorders e.g. corneal infection. The polypeptides can also
CC be used to aid wound healing and epithelial cell proliferation, to
CC prevent skin aging due to sunburn, to maintain organs before
CC transplantation, for supporting cell culture of primary tissues, to
CC regenerate tissues and in chemotaxis. The polypeptides can also be used
CC as a food additive or preservative to increase or decrease storage
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
CC minerals, cofactors and other nutritional components. Numerous
CC examples of diseases and disorders treated by the nucleic acids and
CC proteins are given in the specification. The present sequence

Query Match 65.3%; Score 32; DB 22; Length 63;
Best Local Similarity 55.6%; Pred. NO. 49;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db 11 RLNEYSVNL 19

RESULT 21
AAW22495

ID AAW22495 standard; Protein; 121 AA.
XX
AC AAW22495;
XX
DT 10-MAR-1998 (first entry)
XX
DE Phaffia derived glyceraldehyde-3-phosphate dehydrogenase PRCDNA87.
XX
KW Glyceraldehyde-3-phosphate dehydrogenase; GAPDH; carotenoid;
KW synthesis; promoter; recombinant DNA; astaxanthin; ribosomal protein;
KW food colouring.
XX
OS Phaffia rhodozyma.
XX
PN WO9723633-A1.
XX
PD 03-JUL-1997.
XX
PF 23-DEC-1996; 96WO-EP05887.
XX
PR 11-APR-1996; 96EP-0200943.
PR 22-DEC-1995; 95EP-0203620.
XX
XX (KONN) GIST-BROCADES BV.
PA (OOLJ/) OOLJEN A J J.
XX
XX Verdoes JC, Wery J;
PI
XX WPI; 1997-351068/32.
DR N-PSDB; AAT72939.
XX
XX Phaffia derived GAPDH and carotenoid synthesis genes and promoter
PT fragment - used in the recombinant production of therapeutically
PT useful proteins e.g. carotenoids for use in food colouring
XX
XX Claim 5; Page 91-92; 118pp; English.

CC The present sequence represents a Phaffia derived glyceraldehyde-3-
CC phosphate dehydrogenase (GAPDH). The nucleic acid encoding this protein
CC can be used in the novel recombinant DNA of the present invention. The
CC recombinant DNA comprises a transcription promoter operably linked to a
CC downstream sequence to be expressed, where the transcription promoter
CC comprises a region found upstream of the open reading frame (ORF) of a
CC highly expressed Phaffia gene (preferably GAPDH, a ribosomal protein
CC gene, an enzyme involved in the biosynthesis pathway). The recombinant
CC DNA can be used to transform hosts, preferably Phaffia. These
CC transformed hosts are then used in the recombinant production of GAPDH
CC or an enzyme involved in carotenoid synthesis, preferably astaxanthin.
CC They may also be used to produce a pharmaceutical product. Purified
CC carotenoids can be used as colourants in food and/or feed, and also in
CC cosmetics.

XX
SQ Sequence 121 AA;

Query Match 65.3%; Score 32; DB 18; Length 121;
Best Local Similarity 83.3%; Pred. NO. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9
Db 19 EYTIHL 24

RESULT 22
AAG14213

ID AAG14213 standard; Protein; 276 AA.
XX
AC AAG14213;
XX
DT 17-OCT-2000 (first entry)
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 13988.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0123788.
PR 29-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.

```
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157253.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 65.3%; Score 32; DB 21; Length 276;
Best Local Similarity 55.6%; Pred. No. 2.6e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
    :|||: |||
Db 40 EMNEFIMHL 48

RESULT 23
AAG47496
ID AAG47496 standard; Protein; 284 AA.
XX AC AAG47496;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 59871.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
OS
```

```
XX EP1033405-A2.
XX 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
```

```
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.

      Query Match      65.3%; Score 32; DB 21; Length 284;
      Best Local Similarity 55.6%; Pred. No. 2.7e+02;
      Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
   :|||: ||
Db 40 EMNEFINHL 48

RESULT 24
AAG14212
ID AAG14212 standard; Protein; 316 AA.
XX
AC AAG14212;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 13987.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
```

PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139859.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.

PR	04-OCT-1999;	99US-0157117.	PR	06-APR-1999;	99US-0128234.
PR	05-OCT-1999;	99US-0157753.	PR	08-APR-1999;	99US-0128714.
PR	06-OCT-1999;	99US-0157865.	PR	16-APR-1999;	99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	28-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	13-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159330.	PR	04-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999;	99US-0132486.
PR	14-OCT-1999;	99US-0159638.	PR	07-MAY-1999;	99US-0132487.
PR	18-OCT-1999;	99US-0159584.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160741.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160767.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134370.
PR	21-OCT-1999;	99US-0160815.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160980.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160981.	PR	20-MAY-1999;	99US-0135124.
PR	22-OCT-1999;	99US-0160989.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161404.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161405.	PR	25-MAY-1999;	99US-0136021.
PR	25-OCT-1999;	99US-0161406.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161359.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161360.	PR	01-JUN-1999;	99US-0137222.
PR	26-OCT-1999;	99US-0161361.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161920.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161992.	PR	07-JUN-1999;	99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999;	99US-0138094.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138540.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138847.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUN-1999;	99US-0139119.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	17-JUN-1999;	99US-0139453.
PR	29-OCT-1999;	99US-0162142.	PR	17-JUN-1999;	99US-0139453.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139454.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139455.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139456.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139457.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139457.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139458.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139459.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139460.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139461.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139462.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139463.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139750.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139763.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUN-1999;	99US-0139899.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140353.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140354.
PR	29-OCT-1999;	99US-0162142.	PR	24-JUN-1999;	99US-0140695.
PR	29-OCT-1999;	99US-0162142.	PR	28-JUN-1999;	99US-0140823.
PR	29-OCT-1999;	99US-0162142.	PR	29-JUN-1999;	99US-0140931.
PR	29-OCT-1999;	99US-0162142.	PR	30-JUN-1999;	99US-0141287.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0141842.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0142154.
PR	29-OCT-1999;	99US-0162142.	PR	02-JUL-1999;	99US-0142055.
PR	29-OCT-1999;	99US-0162142.	PR	06-JUL-1999;	99US-0142390.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUL-1999;	99US-0142803.
PR	29-OCT-1999;	99US-0162142.	PR	09-JUL-1999;	99US-0142920.
PR	29-OCT-1999;	99US-0162142.	PR	12-JUL-1999;	99US-0142977.
PR	29-OCT-1999;	99US-0162142.	PR	13-JUL-1999;	99US-0143542.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUL-1999;	99US-0143624.
PR	29-OCT-1999;	99US-0162142.	PR	15-JUL-1999;	99US-0144005.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUL-1999;	99US-0144085.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUL-1999;	99US-0144086.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144325.
PR	29-OCT-1999;	99US-0162142.	PR	19-JUL-1999;	99US-0144331.

Query Match 65.3%; Score 32; DB 21; Length 316;

Best Local Similarity 55.6%; Pred. No. 3e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

:|:|:|:|:

Db 80 EMNEFIMHL 88

RESULT 25

AAG47495

ID AAG47495 standard; Protein; 324 AA.

XX AC AAG47495;

XX DT 18-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 59870.

XX KW Protein identification; signal transduction pathway; metabolic pathway;

XX KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0146264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 19-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145099.
PR 23-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.

PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 65.3%; Score 32; DB 21; Length 324;
Best Local Similarity 55.6%; Pred. No. 3.1e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KMNEVTVHL 9
Db 80 EMNEFIMHL 88
:|||: |||

RESULT 26

AAG14211
ID AAG14211 standard; Protein; 364 AA.

XX
AC AAG14211;

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 13986.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

OS Arabidopsis thaliana.

PN EP1033405-A2.

XX PD 06-SEP-2000.

PF 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 9905-0121825.

PR 05-MAR-1999; 9905-0123180.

PR 09-MAR-1999; 9905-0123548.

PR 23-MAR-1999; 9905-0125788.

PR 25-MAR-1999; 9905-0126264.

PR 29-MAR-1999; 9905-0126785.

PR 01-APR-1999; 9905-0127462.

PR 06-APR-1999; 9905-0128234.

PR 08-APR-1999; 9905-0128714.

PR 16-APR-1999; 9905-0128845.

PR 19-APR-1999; 9905-0130077.

PR 21-APR-1999; 9905-0130449.

PR 23-APR-1999; 9905-0130510.

PR 23-APR-1999; 9905-0130891.

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135355.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150584.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.


```
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match          65.3%; Score 32; DB 21; Length 364;
Best Local Similarity 55.6%; Pred. No. 3.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db 128 ENNEFIMHL 136

RESULT 27
AAG47494
ID AAG47494 standard; Protein; 372 AA.
AC AAG47494;
XX
XX 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 59869.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 28-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
```

```
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147933.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155133.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160771.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 65.3%; Score 32; DB 21; Length 372;
Best Local Similarity 55.6%; Pred. No. 3.6e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 KMNEYTVHL 9
Db 128 EMNEFIMHL 136
:|:|:|:|:|

RESULT 28
AAU33973
ID AAU33973 standard; Protein; 398 AA.
AC AAU33973;
XX 14-FEB-2002 (first entry)
DE Staphylococcus aureus cellular proliferation protein #249.
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX Staphylococcus aureus.
XX WO200170955-A2.
XX 27-SEP-2001.
XX 21-MAR-2001; 2001WO-US09180.
XX 21-MAR-2000; 2000US-191078P.
PR 23-MAY-2000; 2000US-206848P.
PR 26-MAY-2000; 2000US-207727P.
PR 23-OCT-2000; 2000US-242578P.
PR 27-NOV-2000; 2000US-253625P.
PR 22-DEC-2000; 2000US-257931P.
PR 16-FEB-2001; 2001US-269308P.
XX (ELIT-) ELITRA PHARM INC.
XX.
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
XX WPI: 2001-611495/70.
DR N-PSDB; AAS1832.
XX
New polynucleotides for the identification and development of
antibiotics, comprise sequences of antisense nucleic acids -
Example 3; Seq ID No 5469; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
```

CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX SQ Sequence 398 AA;

Query Match 65.3%; Score 32; DB 22; Length 398;
 Best Local Similarity 62.5%; Pred. No. 3.9e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
 :||| |
 Db 367 KLNETHYTH 374

RESULT 29
 AAU36825
 ID AAU36825 standard; Protein; 409 AA.
 XX
 AC AAU36825;
 XX

DT 14-FEB-2002 (first entry).

DE Staphylococcus aureus cellular proliferation protein #995.

KW Antisense; prokaryotic cellular proliferation protein;
 KW antibiotic; antibacterial; drug design.

XX Staphylococcus aureus.

OS WO200170955-A2.

XX 27-SEP-2001.

PF 21-MAR-2001; 2001WO-US09180.

XX 21-MAR-2000; 2000US-191078P.

PR 23-MAY-2000; 2000US-206848P.

PR 26-MAY-2000; 2000US-207727P.

PR 23-OCT-2000; 2000US-242578P.

PR 27-NOV-2000; 2000US-253625P.

PR 22-DEC-2000; 2000US-257931P.

PR 16-FEB-2001; 2001US-269308P.

XX (ELIT-) ELITRA PHARM INC.

XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;

PI Yamamoto RT, Xu HH;

XX WPI: 2001-611495/70.

DR N-PSDB; AA554684.

XX New polynucleotides for the identification and development of

PT antibiotics, comprise sequences of antisense nucleic acids -

PS Example 3; Seq ID No 12418; 511pp; English.

XX The invention relates to antisense inhibitors of genes essential to

CC prokaryotic cellular proliferation, their use in identifying the

CC genes, their use in the discovery of novel antibiotics, the essential

CC genes themselves and the encoded proteins. The prokaryotes used are

CC to identify proteins used in proliferation, to express these proteins,
 CC and to obtain antibodies capable of binding to the expressed proteins.
 CC The proteins can be used to screen compounds in rational drug discovery
 CC programmes. The antisense nucleic acid sequence is also useful to screen
 CC for homologous nucleic acids which are required for cell proliferation in
 CC a wide variety of organisms. The present sequence represents an
 CC essential prokaryotic cellular proliferation protein.
 CC Note: The sequence data for this patent did not form part
 CC of the printed specification, but was obtained in electronic
 CC format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX SQ Sequence 409 AA;

Query Match 65.3%; Score 32; DB 22; Length 409;
 Best Local Similarity 62.5%; Pred. No. 4e+02;
 Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
 :||| |
 Db 378 KLNETHYTH 385

RESULT 30
 AAP80147
 ID AAP80147 standard; protein; 639 AA.

XX AAP80147;
 AC

XX 11-DEC-1990 (first entry)

DE Tetracycline resistance protein from the tetM:Tn 1545 gene.

KW tetM gene; tetracycline resistance protein; probes; transposon Tn1545.

XX Streptococcus faecalis BM 4127.

XX FR2602794-A.

XX 19-FEB-1988.

XX 28-JUL-1986; 86FR-0010921.

XX 28-JUL-1986; 86FR-0010921.

XX (INSP) INST PASTEUR.

XX (CNRS) CENT NAT RECH SCI.

XX WPI: 1988-100268/15.

XX N-PSDB; AA80169.

XX New DNA fragment contg. tetM gene - for coding new tetracycline

PT resistance protein, useful for making probes to detect such

PT resistance.

XX Claim 6; Page 22; 29pp; French.

XX The hydrophilic protein has a mol. wt. of 72500 daltons and

CC can be used to make specific antisera.

XX Sequence 639 AA;

Query Match 65.3%; Score 32; DB 9; Length 639;

Best Local Similarity 55.6%; Pred. No. 6.6e+02;

Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVH 9
 :||| |
 Db 426 KKAETHYTHI 434

RESULT 31
 ABB65057

ID ABB65057 standard; Protein; 678 AA.
XX AC ABB65057;
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 21963.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX KW pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX DR WPI: 2001-656860/75.
XX DR N-PSDB; ABL09160.
XX XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX PS Disclosure; SEQ ID NO 21963; 21pp + Sequence Listing; English.
XX CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 678 AA;
Query Match 65.3%; Score 32; DB 22; Length 678;
Best Local Similarity 75.0%; Pred. No. 7.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 MNEYTVHL 9
DB 422 LNESTVHL 429
RESULT 32
ABG06295
ID ABG06295 standard; Protein; 683 AA.
XX AC ABG06295;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #6286.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX XX

PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX XX Drmanac RT, Liu C, Tang YT;
XX PI WPI: 2001-639362/73.
XX DR N-PSDB; AAS70482.
XX XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX XX Claim 20; SEQ ID NO 36654; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 683 AA;
Query Match 65.3%; Score 32; DB 22; Length 683;
Best Local Similarity 62.5%; Pred. No. 7.2e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
DB 152 KLNHTVH 159
RESULT 33
AAM39237
ID AAM39237 standard; Protein; 688 AA.
XX AC AAM39237;
XX DT 22-OCT-2001 (first entry)
XX DE Human polypeptide SEQ ID NO 2382.
XX KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW ankyrotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX XX

OS Homo sapiens.
XX WO200153312-A1.
XX
XX
XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
XX 25-APR-2000; 2000US-0552317.
XX 09-JUL-2000; 2000US-0598042.
XX 19-JUL-2000; 2000US-0620312.
XX 03-AUG-2000; 2000US-0653450.
XX 14-SEP-2000; 2000US-0662191.
XX 19-OCT-2000; 2000US-0693036.
XX 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
XX N-PSDB; AAI58393.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
XX such as central nervous system injuries .
XX
XX Example 4; SEQ ID NO 2382; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
XX the encoded polypeptides (AAM38642-AAM42213) with nootropic,
XX immunosuppressant and cytostatic activity. The polynucleotides are useful
XX in gene therapy. A composition containing a polypeptide or polynucleotide
XX of the invention may be used to treat diseases of the peripheral nervous
XX system, such as peripheral nervous injuries, peripheral neuropathy and
XX localised neuropathies and central nervous system diseases, such as
XX Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
XX lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
XX utilisation of the activities such as: Immune system suppression,
XX Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX and thrombolytic activity, cancer diagnosis and therapy, drug screening,
XX assays for receptor activity, arthritis and inflammation, leukaemias and
XX C.N.S disorders.
XX Note: The sequence data for this patent did not form part of the printed
XX specification.
XX
XX Sequence 688 AA;
XX
XX Query Match 65.3%; Score 32; DB 22; Length 688;
XX Best Local Similarity 71.4%; Pred. No. 7.2e+02;
XX Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 MNEYTVH 8
XX |||::||
XX Db 188 MNEFSVH 194
XX
XX RESULT 34
XX AAB65597
XX ID AAB65597 standard; Protein; 688 AA.
XX
XX AAB65597;
XX
XX 27-MAR-2001 (first entry)
XX
XX Novel protein kinase, SEQ ID NO: 122.
XX
XX Human; mouse; protein kinase; antiarthritic; antisclerotic; osteopathic;
XX immunosuppressive; cardiant; renal; antiinflammatory; antiasthmatic;
XX dermatological; antidiabetic; antinfertility; gene therapy; vaccine;
XX immune disorder; cardiovascular disease; neurodegenerative disease;

KW cancer; autoimmune disorder; stroke; inflammatory bowel disease;
KW inflammatory pelvic disease; multiple sclerosis; psoriasis.
XX
XX Homo sapiens.
XX
XX WO200073469-A2.
XX
XX 07-DEC-2000.
XX
XX 26-MAY-2000; 2000WO-US14842.
XX
XX 28-MAY-1999; 99US-0136503.
XX (SUGE-) SUGEN INC.
XX
XX Plowman GD, Martinez R, Whyte D, Sudersanam S;
XX
XX WPI: 2001-032161/04.
XX N-PSDB; AAF44622.
XX
XX Nucleic acids encoding kinase polypeptides, useful for diagnosing and
XX treating immune-related diseases and disorders, cardiovascular disease,
XX neurodegenerative diseases and/or cancers .
XX
XX Claim 10; Fig 1; 310pp; English.
XX
XX The present sequence is a novel protein kinase. The novel protein kinases
XX and the nucleic acids that encode them may be used in the treatment and
XX diagnosis of diseases associated with inappropriate kinase expression
XX such as immune-related diseases and disorders, cardiovascular disease,
XX neurodegenerative diseases and/or cancers. The nucleic acids and
XX complementary sequences may also be used as DNA probes in diagnostic
XX assays. The kinase polypeptides may be used as antigens in the production
XX of antibodies of kinase expression and activity. Anti-kinase antibodies
XX and kinase antagonists may also be used to down regulate kinase
XX expression and activity. Diseases related to kinase expression and
XX activity include rheumatoid arthritis, atherosclerosis, autoimmune
XX disorders, complications of organ transplantation, myocardial infarction,
XX immune disorders, cardiomyopathies, strokes, renal failure,
XX oxidative-stress related disorders, chronic inflammatory bowel disease,
XX chronic inflammatory pelvic disease, multiple sclerosis, asthma,
XX osteoarthritis, psoriasis, rhinitis, autoimmunity, diabetes, cancers and
XX reproductive disorders.
XX
XX Sequence 688 AA;
XX
XX Query Match 65.3%; Score 32; DB 22; Length 688;
XX Best Local Similarity 71.4%; Pred. No. 7.2e+02;
XX Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 MNEYTVH 8
XX |||::||
XX Db 188 MNEFSVH 194
XX
XX RESULT 35
XX ABG08784
XX ID ABG08784 standard; Protein; 692 AA.
XX
XX ABG08784;
XX
XX 13-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #8775.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR N-PSDB; AAS72971.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics; gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 39143; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences; (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 692 AA:
SQ
Query Match 65.3%; Score 32; DB 22; Length 692;
Best Local Similarity 62.5%; Pred. No. 7.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMEYTVH 8
|:|:|:|
DB 161 KLNHHTVH 168
RESULT 36
AAM41023
ID AAM41023 standard; Protein; 698 AA.
XX
XX AAM41023;
XX
XX 22-OCT-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 5954.
XX
XX Human; nontropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW Leukaemia.
XX
XX Homo sapiens.
OS
XX
XX WO200153312-A1.
PN

XX 26-JUL-2001.
XX
XX 26-DEC-2000; 2000WO-US34263.
XX
XX 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR N-PSDB; AAI60179.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 5954; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AA42213) with nontropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
XX Sequence 698 AA;
SQ
Query Match 65.3%; Score 32; DB 22; Length 698;
Best Local Similarity 71.4%; Pred. No. 7.3e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 MNEYTVH 8
|:|:|:|
DB 198 MNEFSVH 204
RESULT 37
AAU37828
ID AAU37828 standard; Protein; 752 AA.
XX
XX AAU37828;
XX
XX 14-FEB-2002 (first entry)
XX
XX Streptococcus pneumoniae cellular proliferation protein #257.
DE
XX Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
KW
XX Streptococcus pneumoniae.
OS
XX
XX WO200170955-A2.
PN
XX

```

PD XX 27-SEP-2001.
PF PF 21-MAR-2001; 2001WO-US09180.
XX XX
PR XX 21-MAR-2000; 2000US-191078P.
PR XX 23-MAY-2000; 2000US-206848P.
PR XX 26-MAY-2000; 2000US-207727P.
PR XX 23-OCT-2000; 2000US-242578P.
PR XX 27-NOV-2000; 2000US-253625P.
PR XX 22-DEC-2000; 2000US-257931P.
PR XX 16-FEB-2001; 2001US-269308P.
XX XX
PA (ELIT-) ELITRA PHARM INC.
XX XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS55687.
XX XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX XX
XX Example 3; Seq ID No 13421; 51pp; English.
XX XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 752 AA;

Query Match 65.3%; Score 32; DB 22; Length 752;
Best Local Similarity 55.6%; Pred. NO. 8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db |:|:|:|:|
7 KINDSTIHL 15

RESULT 38
AAU38068
ID AAU38068 standard; Protein; 752 AA.
XX XX
AC AAU38068;
XX XX
DT 14-FEB-2002 (first entry)
XX XX
DE Streptococcus pneumoniae cellular proliferation protein #497.
XX XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX XX
OS Streptococcus pneumoniae.
XX XX
PN WO200170955-A2.
XX XX

PD XX 27-SEP-2001.
PF PF 21-MAR-2001; 2001WO-US09180.
XX XX
PR XX 21-MAR-2000; 2000US-191078P.
PR XX 23-MAY-2000; 2000US-206848P.
PR XX 26-MAY-2000; 2000US-207727P.
PR XX 23-OCT-2000; 2000US-242578P.
PR XX 27-NOV-2000; 2000US-253625P.
PR XX 22-DEC-2000; 2000US-257931P.
PR XX 16-FEB-2001; 2001US-269308P.
XX XX
PA (ELIT-) ELITRA PHARM INC.
XX XX
XX Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX XX
DR WPI; 2001-611495/70.
DR N-PSDB; AAS55687.
XX XX
XX New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX XX
XX Example 3; Seq ID No 13421; 51pp; English.
XX XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC Escherichia coli, Staphylococcus aureus, Salmonella typhi, Klebsiella
CC pneumoniae, Pseudomonas aeruginosa and Enterococcus faecalis. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an
CC essential prokaryotic cellular proliferation protein.
CC Note: The sequence data for this patent did not form part
CC of the printed specification, but was obtained in electronic
CC format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX XX
SQ Sequence 752 AA;

Query Match 65.3%; Score 32; DB 22; Length 752;
Best Local Similarity 55.6%; Pred. NO. 8e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db |:|:|:|:|
7 KINDSTIHL 15

RESULT 39
ABG01448
ID ABG01448 standard; Protein; 796 AA.
XX XX
AC ABG01448;
XX XX
DT 13-FEB-2002 (first entry)
XX XX
DE Novel human diagnostic protein #1439.
XX XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX XX
OS Homo sapiens.
XX XX
PN WO200175067-A2.
XX XX

```

PD 11-OCT-2001.
XX 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS65635.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 31807; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 796 AA;
Query Match 65.3%; Score 32; DB 22; Length 796;
Best Local Similarity 62.5%; Pred. No. 8.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
I:|:|
Db 152 KLNHHTVH 159
RESULT 40
ABG11797
ID ABG11797 standard; Protein: 796 AA.
XX AC ABG11797;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #11788.
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX WO200175067-A2.
XX PN 11-OCT-2001.
XX PD 11-OCT-2001.
XX PF 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS75984.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 42156; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX Note: The sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
SQ Sequence 796 AA;
Query Match 65.3%; Score 32; DB 22; Length 796;
Best Local Similarity 62.5%; Pred. No. 8.5e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
I:|:|
Db 152 KLNHHTVH 159
RESULT 41
ABG05821
ID ABG05821 standard; Protein: 882 AA.
XX AC ABG05821;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #5812.
XX Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX WO200175067-A2.
XX PN 11-OCT-2001.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX

PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PA (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
PI WPI; 2001-639362/73.
XX N-PSDB; AAS70008.
DR New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 36180; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 882 AA;
Query Match 65.3%; Score 32; DB 22; Length 882;
Best Local Similarity 62.5%; Pred. NO. 9.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
I:I :|||
Db 152 KLNHNTVH 159
RESULT 42
ABG06296
ID ABG06296 standard; Protein: 917 AA.
XX AC ABG06296;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #6287.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX N-PSDB; AAS70483.
DR New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID No 36655; 103pp; English.
PS The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 917 AA;
Query Match 65.3%; Score 32; DB 22; Length 917;
Best Local Similarity 62.5%; Pred. NO. 1e+03;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
I:I :|||
Db 109 KLNHNTVH 116
RESULT 43
ABG08780
ID ABG08780 standard; Protein: 917 AA.
XX AC ABG08780;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #8771.
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.

```
XX Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
DR N-PSDB; AAS72967.
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX Claim 20; SEQ ID NO 39139; 103pp; English.
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX Sequence 917 AA;
SQ Query Match 65.3%; Score 32; DB 22; Length 917;
Best Local Similarity 62.5%; Pred. NO. 1e+03;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
DB 109 KLNHNTVH 116
RESULT 44
AAM79950
ID AAM79950 standard; Protein; 951 AA.
XX AAM79950;
AC AAM79950;
DT 06-NOV-2001 (first entry)
XX Human protein SEQ ID NO 3596.
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX Homo sapiens.
XX WO200157190-A2.
PN 09-AUG-2001.
PD 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US04098.
PF 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX (HYSE-) HYSEQ INC.
PA Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
XX Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX WPI; 2001-476283/51.
DR N-PSDB; AAK53083.
DR Nucleic acids encoding polypeptides with cytokine-like activities,
XX useful in diagnosis and gene therapy -
PT Claim 20; Page 393; 6221pp; English.
XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX Sequence 951 AA;
SQ Query Match 65.3%; Score 32; DB 22; Length 951;
Best Local Similarity 62.5%; Pred. NO. 1e+03;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
DB 161 KLNHNTVH 168
RESULT 45
ABG08785
ID ABG08785 standard; Protein; 1015 AA.
XX ABG08785;
AC ABG08785;
DT 13-FEB-2002 (first entry)
XX Novel human diagnostic protein #8776.
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX Homo sapiens.
XX WO200175067-A2.
PN 11-OCT-2001.
PD 30-MAR-2001; 2001WO-US08631.
PF 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX (HYSE-) HYSEQ INC.
PA Drmanac RT, Liu C, Tang YT;
PI
```

DR WPI: 2001-639362/73.
 DR N-PSDB; AAS72972.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 39144; 103pp; English.
 PS
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: the sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 1015 AA;

Query Match 65.3%; Score 32; DB 22; Length 1015;
 Best Local Similarity 62.5%; Pred. NO. 1.1e+03;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
 Db 109 KLNHFTVH 116
 !:!:|||

Search completed: November 6, 2002, 12:05:00
 Job time : 26.7778 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 18.4444 Seconds
(without alignments)
84.413 Million cell updates/sec

Title: US-09-905-083-31
Perfect score: 49
Sequence: 1 KMNEYTVHL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	39	79.6	210	9 Q9G0B3	Q9G0B3 bacteriophage
2	37	75.5	393	17 Q9V291	Q9V291 pyrococcus
3	36	73.5	627	5 Q9XU04	Q9XU04 caenorhabd
4	36	73.5	635	10 Q9LQW1	Q9LQW1 arabidopsis
5	36	73.5	1829	5 Q19815	Q19815 caenorhabd
6	35	71.4	159	16 Q92711	Q92711 listeria in
7	35	71.4	433	4 Q9X315	Q9X315 homo sapien
8	35	71.4	505	16 Q9RV79	Q9RV79 delnococtus
9	35	71.4	555	16 Q92621	Q92621 listeria in
10	35	71.4	667	4 Q15344	Q15344 homo sapien
11	35	71.4	667	11 P82457	P82457 mus spretus
12	35	71.4	667	11 P82458	P82458 rattus norv
13	35	71.4	667	13 Q90WD1	Q90WD1 gallus gall
14	35	71.4	680	11 Q70583	Q70583 mus musculu
15	35	71.4	685	4 Q9UJV3	Q9UJV3 homo sapien
16	35	71.4	685	11 Q9QU56	Q9QU56 mus musculu

17	35	71.4	715	4 Q9UJR9	Q9UJR9 homo sapien
18	34	69.4	270	11 Q99KL9	Q99KL9 mus musculu
19	34	69.4	421	16 Q92ES2	Q92ES2 listeria in
20	34	69.4	1946	5 Q97291	Q97291 plasmodium
21	33	67.3	218	16 Q9KF88	Q9KF88 bacillus ha
22	33	67.3	227	16 Q9KDN9	Q9KDN9 bacillus ha
23	33	67.3	327	3 Q96V06	Q96V06 cryptonec
24	33	67.3	422	3 Q74547	Q74547 schizosacch
25	33	67.3	525	16 Q9K8I0	Q9K8I0 bacillus ha
26	33	67.3	664	16 Q9WY08	Q9WY08 thermotoga
27	33	67.3	677	16 Q9A7E2	Q9A7E2 caulobacter
28	33	67.3	776	4 Q96CC7	Q96CC7 homo sapien
29	33	67.3	799	11 Q99K69	Q99K69 mus musculu
30	33	67.3	1308	5 Q96129	Q96129 plasmodium
31	32	65.3	47	8 Q95CP1	Q95CP1 jacksonia h
32	32	65.3	49	16 Q99UE1	Q99UE1 staphylococ
33	32	65.3	130	16 Q97F09	Q97F09 clostridium
34	32	65.3	133	12 Q91DR7	Q91DR7 sapporo vir
35	32	65.3	133	12 Q91DP2	Q91DP2 sapporo vir
36	32	65.3	133	12 Q91DP1	Q91DP1 sapporo vir
37	32	65.3	133	12 Q91DN8	Q91DN8 sapporo vir
38	32	65.3	133	12 Q91DN6	Q91DN6 sapporo vir
39	32	65.3	134	12 Q91DR6	Q91DR6 sapporo vir
40	32	65.3	134	12 Q91DR5	Q91DR5 sapporo vir
41	32	65.3	134	12 Q91DP6	Q91DP6 sapporo vir
42	32	65.3	134	12 Q91DN4	Q91DN4 sapporo vir
43	32	65.3	144	17 Q57714	Q57714 pyrococcus
44	32	65.3	181	5 Q9GZ14	Q9GZ14 caenorhabd
45	32	65.3	192	16 Q9PMC4	Q9PMC4 campylobact

ALIGNMENTS

RESULT 1

Q9G0B3 PRELIMINARY; PRT; 210 AA.
 ID Q9G0B3
 AC Q9G0B3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE DISTAL TAIL FIBER LOCUS, PARTIAL SEQUENCE (FRAGMENT).
 GN T.
 OS Bacteriophage ARL.
 OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Myoviridae;
 OC T4-like phages.
 OX NCBI_TaxID=66711;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20485545; PubMed=11029414;
 RA Yu S.L., Ko K.L., Chen C.S., Chang Y.C., Syu W.J.;
 RT "Characterization of the Distal Tail Fiber Locus and Determination of
 RT the Receptor for Phage ARL, Which Specifically Infects Escherichia
 RT coli O157:H7.";
 RL J. Bacteriol. 182:5962-5968(2000).
 DR EMBL: AF208841; AAG29756.1; -
 FT NON_TER 210
 SQ SEQUENCE 210 AA; 24348 MW; C6FF585F9FE0DF68 CRC64;

Query Match 79.6%; Score 39; DB 9; Length 210;
 Best Local Similarity 87.5%; Pred. No. 4.7;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
 |:|||||
 Db 139 MDEYTVHL 146

RESULT 2

Q9V291 PRELIMINARY; PRT; 393 AA.
 ID Q9V291
 AC Q9V291;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DE 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE HYPOTHETICAL 45.5 KDA PROTEIN.
 GN PAB2235.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ORSAY;
 RA Heilig R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ248283; CAB49107.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 393 AA; 45485 MW; 9448642FBB3FBF43 CRC64;

 Query Match 75.5%; Score 37; DB 17; Length 393;
 Best Local Similarity 66.7%; Pred. No. 23;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 QY 1 KMNEYTVHL 9
 DB 129 KLPEYTIHL 137

 RESULT 3
 Q9XU04
 ID Q9XU04 PRELIMINARY; PRT; 627 AA.
 AC Q9XU04;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE T28A8.6 PROTEIN.
 GN T28A8.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lloyd C.R.;
 RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; Z92813; CAB07289.1; -
 SQ SEQUENCE 627 AA; 71794 MW; 4A56E4D3658EC1CC CRC64;

 Query Match 73.5%; Score 36; DB 5; Length 627;
 Best Local Similarity 66.7%; Pred. No. 58;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

 QY 1 KMNEYTVHL 9
 DB 62 KMNEYSIEL 70

 RESULT 4
 Q9LQW1
 ID Q9LQW1 PRELIMINARY; PRT; 635 AA.
 AC Q9LQW1;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE F1086.9.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OX eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Chao Q., Shinn P., Dunn P., Buehler E., Kahn S., Kim C., Walker M.,
 RA Williams S., Altafi H., Araujo R., Conn L., Conway A.B., Gonzalez A.,
 RA Hansen N.F., Huizar L., Kremenetskaia I., Lenz C., Li J., Liu S.,
 RA Luros S., Rowley D., Schwartz J., Toriumi M., Vysotskaia V., Yu G.,
 RA Davis R.W., Federspiel N.A., Theologis A., Ecker J.R.;
 RT "Genomic sequence for Arabidopsis thaliana BAC F10B6 from chromosome I."
 RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Ecker J.R.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,
 RA Khan S., Kim C., Altafi H., Bei B., Chin C., Chiou J., Choi E.,
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,
 RA Lee J., Lenz C., Li J., Liu A., Liu J., Liu S., Mukharsky N.,
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,
 RA Theologis A., Ecker J.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AC006917; AAF79248.1; -
 SQ SEQUENCE 635 AA; 72664 MW; 2871C7DA1E4DA65E CRC64;

 Query Match 73.5%; Score 36; DB 10; Length 635;
 Best Local Similarity 66.7%; Pred. No. 58;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

 QY 1 KMNEYTVHL 9
 DB 169 KLENYQTHL 177

 RESULT 5
 Q19815
 ID Q19815 PRELIMINARY; PRT; 1829 AA.
 AC Q19815;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE F26F12.7 PROTEIN (LET-418).
 GN F26F12.7 OR LET-418.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL NZ;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkison-Sproat J., Wohlman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C. elegans."
 RL Nature 368:32-38(1994).

RA SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Wilson R., Bentley D., Gattung S.;
RT "The sequence of *C. elegans* cosmid F26F12.";
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
[3]
RA SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.
[4]
RA SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
[5]
RA SEQUENCE FROM N.A.
RP von Zellewsky T., Palladino F., Brunschwig K., Hajnal A., Mueller F.;
RT "The *C. elegans* MI-2 chromatin-remodeling proteins function in vulval
cell fate determination.";
RL Development 0:0-0(2000).
DR EMBL: U55373; AAC25894.1; --
DR EMBL: AF308445; AAG29838.1; --
DR InterPro: IPR000953; Chromo.
DR InterPro: IPR001410; DEAD.
DR InterPro: IPR002464; DEAH_ATP_helicase.
DR InterPro: IPR001650; Helicase_C.
DR InterPro: IPR001965; PHD.
DR InterPro: IPR000330; SNF2_N.
DR InterPro: IPR001841; znf_ring.
DR Pfam: PF00385; chromo; 1.
DR Pfam: PF00271; helicase_C; 1.
DR Pfam: PF00628; PHD; 2.
DR Pfam: PF00176; SNF2_N; 1.
DR SMART: SM00298; CHROMO; 2.
DR SMART: SM00487; DEXDC; 1.
DR SMART: SM00490; HELIC; 1.
DR SMART: SM00249; PHD; 2.
DR SMART: SM00184; RING; 2.
DR PROSITE: PS50013; CHROMO_2; 2.
DR PROSITE: PS00690; DEAH_ATP_HELICASE; UNKNOWN_1.
KW ATP-binding; Helicase.
SQ SEQUENCE 1829 AA; 209150 MW; 1A887E990C63B661 CRC64;

Query Match 73.5%; Score 36; DB 5; Length 1829;
Best Local Similarity 71.4%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVH 8
| | | | |
DB 765 LNEYTH 771

RESULT 6
Q927II PRELIMINARY; PRT; 159 AA.
AC Q927II;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LIN2807 PROTEIN.
GN LIN2807.
OS *Listeria innocua*.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC *Bacillus/Staphylococcus* group; *Listeria*.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CLIP 11262 / SEROVAR 6A;
RX PubMed-11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,

RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kreft J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordstiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Rammel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of *Listeria* species.";
RL Science 294:849-852(2001).
DR EMBL: AL596173; CAC98033.1; --
DR ListList; LIN02807; --
KW Complete proteome.
SQ SEQUENCE 159 AA; 18259 MW; 70A7217BF6B12FDA CRC64;

Query Match 71.4%; Score 35; DB 16; Length 159;
Best Local Similarity 75.0%; Pred. No. 23;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
| | | | |
DB 139 KMSEWTVH 146

RESULT 7
Q9Y3I5 PRELIMINARY; PRT; 433 AA.
AC Q9Y3I5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE DAL19P20.2 (NOVEL FIBRONECTIN TYPE III DOMAIN CONTAINING PROTEIN
DE SIMILAR TO RING FINGER PROTEIN MIDI (MIDLINE 1)) (FRAGMENT).
GN DAL19P20.2.
OS *Homo sapiens* (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilson S.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL034399; CAB41046.1; --
DR InterPro: IPR003649; Bbox_C.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR001870; Gamma_carboxylase.
DR InterPro: IPR003877; SPRY.
DR InterPro: IPR003878; SPRY_domain.
DR Pfam: PF00041; fn3; 1.
DR Pfam: PF00622; SPRY; 1.
DR SMART: SM00502; BEC; 1.
DR SMART: SM00060; FN3; 1.
DR SMART: SM00449; SPRY; 1.
FT NON_TER 1
SQ SEQUENCE 433 AA; 49294 MW; 6CC779B3DB5C683D CRC64;

Query Match 71.4%; Score 35; DB 4; Length 433;
Best Local Similarity 75.0%; Pred. No. 63;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
| | | | |
DB 192 KQHYTVH 199

RESULT 8
Q9RV79 PRELIMINARY; PRT; 505 AA.
ID Q9RV79;
AC Q9RV79;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE CONSERVED HYPOTHETICAL PROTEIN.
GN DR1150.
OS Deinococcus radiodurans.
OC Bacteria: Thermus/Deinococcus group; Deinococcales, Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL; AB001964; AAF10722.1; -.
DR TIGR; DR1150; -.
DR InterPro; IPR001993; Mitoch_carrier.
DR InterPro; IPR001861; UPF0004.
DR Pfam; PF00919; UPF0004; 1.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR PROSITE; PS01278; UPF0004; 1.
KW Complete: proteome.
SQ SEQUENCE 505 AA; 55752 MW; 884C5BB398D632C2 CRC64;

Query Match 71.4%; Score 35; DB 16; Length 505;
Best Local Similarity 66.7%; Pred. No. 74;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Oy 1 KMNEYTVHL 9
Db 57 QMNEYDTHL 65
:|||||

RESULT 9
Oy 92621 PRELIMINARY; PRT; 555 AA.
ID Q92621
AC Q92621
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE LIN2898 PROTEIN.
GN LIN2898.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusnlok C., Amend A.,
RA Raquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entlan K.-D., Fshih H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kref J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablos B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehlund J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL596173; CAC98124.1; -.
DR ListLit; LIN2898; -.
KW Complete: proteome.
SQ SEQUENCE 555 AA; 63000 MW; 90755353C9E72D42 CRC64;

Query Match 71.4%; Score 35; DB 16; Length 555;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 KMNEYTVHL 9
Db 483 KINEYTDL 491
:|||||

RESULT 10
Oy 15344 PRELIMINARY; PRT; 667 AA.
ID O15344
AC O15344; Q75361; Q9BZX5;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MIDLINE 1 PROTEIN (PUTATIVE TRANSCRIPTION FACTOR XPRF) (TRIPARTITE
DE MOTIF PROTEIN TRIM18).
GN MID1 OR FXY OR XPRF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS OS MET-458 DEL AND FIDSGRHL-534 INS.
RC TISSUE=NEURON;
RX MEDLINE=98016411; PubMed=9354791;
RA Quaderi N., Schweiger S., Gaudenz K., Franco B., Rugarli E.I.,
RA Berger W., Feldman G.J., Volta M., Andolfi G., Gilgenkrantz S.,
RA Marion R.W., Hennekam R.C.M., Opl J.M., Muenke K., Ropers H.H.,
RA Ballabio A.;
RT "Opitz G/BBB syndrome, a defect of midline development, is due to
RT mutations in a new RING finger gene on Xp22.";
RL Nat. Genet. 17:285-291(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RX MEDLINE=98087583; PubMed=9425238;
RA Perry J., Feather S., Smith A., Palmer S., Ashworth A.;
RT "The human FXY gene is located within Xp22.3: Implications for
RT evolution of the mammalian X chromosome.";
RL Hum. Mol. Genet. 7:299-305(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=FETAL KIDNEY;
RX MEDLINE=98390188; PubMed=9722948;
RA Van den Veyver I.B., Cormier T.A., Jurecic V., Baldini A.,
RA Zoghbi H.Y.;
RT "Characterization and physical mapping in human and mouse of a novel
RT RING finger gene in Xp22.";
RL Genomics 51:251-261(1998).
RN [4]
RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX MEDLINE=21231161; PubMed=11331580;
RA Rignand D., Meroni G., Fantozzi A., Merla G., Cairo S., Luzi L.,
RA Riganelli D., Zanaria E., Messali S., Cainarca S., Guffanti A.,
RA Minucci S., Pellicci P.G., Ballabio A.;
RT "The tripartite motif family identifies cell compartments.";
RL EMBO J. 20:2140-2151(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20485428; PubMed=11030761;
RA Cox T.C., Allen L.R., Cox L.L., Hopwood B., Goodwin B., Haan E.,
RT "New mutations in MID1 provide support for loss of function as the
RT cause of X-linked opitz syndrome.";
RL Hum. Mol. Genet. 9:2553-2562(2000).
CC -!- FUNCTION: PUTATIVE TRANSCRIPTIONAL REGULATOR.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA (SHOWN HERE) AND BETA; ARE
CC -!- TISSUE SPECIFICITY: IN THE FETUS, HIGHEST EXPRESSION FOUND IN

CC KIDNEY, FOLLOWED BY BRAIN AND LUNG. EXPRESSED AT LOW LEVELS IN
 CC FETAL LIVER. IN THE ADULT, MOST ABUNDANT IN HEART, PLACENTA AND
 CC BRAIN.
 CC
 CC -!- DISEASE: DEFECTS IN MID1 ARE THE CAUSE OF OPITZ G/BBB SYNDROME
 CC (OS), AN INHERITED DISORDER CHARACTERIZED BY HYPOBLASTIC
 CC GENITAL-URINARY DEFECTS SUCH AS HYPOSPADIAS IN MALES AND SPAYED
 CC LABIA IN FEMALES, LIP-PALATE-LARYNGOTRACHEAL CLEFTS, IMPERFORATE
 CC ANUS, DEVELOPMENTAL DELAY AND CONGENITAL HEART DEFECTS.
 CC
 CC -!- SIMILARITY: CONTAINS 1 C3HC4-CLASS ZINC FINGER.
 CC EMBL: Y13667; CAAT4018.1; -
 DR EMBL: AF035360; AAB99951.1; -
 DR EMBL: AF041206; AAC32998.1; -
 DR EMBL: AF041207; AAC32999.1; -
 DR EMBL: AF041208; AAC33000.1; -
 DR EMBL: AF041209; AAC33001.1; -
 DR EMBL: AF041210; AAC33002.1; -
 DR EMBL: AF230976; AAG50191.1; -
 DR EMBL: AF230977; AAG50192.1; -
 DR EMBL: AF269101; AAG33130.1; -
 DR MIM: 300000; -
 DR InterPro: IPR003649; Bbox_C.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001870; Gamma_carboxylase.
 DR InterPro: IPR003877; SPRY.
 DR InterPro: IPR003878; SPRY_domain.
 DR InterPro: IPR000315; znf_bbox.
 DR InterPro: IPR001841; znf_ring.
 DR Pfam: PF00041; fn3; 1.
 DR Pfam: PF00822; SPRY; 1.
 DR Pfam: PF00843; zf-B_box; 2.
 DR Pfam: PF00097; zf-C3HC4; 1.
 DR PRINTS: PR01406; BBOXZNFINGER.
 DR SMART: SM00502; BBC; 1.
 DR SMART: SM00336; BBOX; 2.
 DR SMART: SM00060; FN3; 1.
 DR SMART: SM00449; SPRY; 1.
 DR SMART: SM00184; RING; 1.
 DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
 DR Coiled coil; Zinc-finger; Metal-binding; Transcription regulation;
 KW Transcription regulation; Nuclear protein; Alternative splicing.
 FT ZN_FING 10 56 C3HC4-TYPE (POTENTIAL).
 FT DOMAIN 170 212 B BOX.
 FT VARSPLIC 205 264 COILED COIL (POTENTIAL).
 FT MISSING 553 667 MISSING (IN ISOFORM BETA).
 FT VARIANT 438 438 MISSING (IN OS).
 FT VARIANT 534 534 V -> VFIDSGRHL (IN OS).
 FT CONFLICT 228 228 T -> P (IN REF. 3; AAC32999).
 FT CONFLICT 484 484 Q -> P (IN REF. 3; AAC32998).
 SQ SEQUENCE 667 AA; 75250 MW; 673C5120018BA619 CRC64;
 Query Match 71.4%; Score 35; DB 4; Length 667;
 Best Local Similarity 75.0%; Pred. No. 98;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KMNHYTVH 8
 DB 444 KQNHVTVH 451
 ID P82457 PRELIMINARY; PRT; 667 AA.
 AC P82457;
 DT 01-JUN-2000 (TrEMBLrel. 14, Created)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last sequence update)
 DE MIDLINE 1 PROTEIN (RING FINGER PROTEIN).
 GN MID1 OR FXV.
 OS Mus spretus (Western wild mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10096;
 RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=99439873; Pubmed=10508587;
 RA Perry J., Ashworth A.;
 RT "Evolutionary rate of a gene affected by chromosomal position.";
 RL Curr. Biol. 9:987-989(1999).
 CC -!- FUNCTION: PUTATIVE TRANSCRIPTIONAL REGULATOR.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
 DR EMBL: AF186460; AAD56246.1; -
 DR MGD: MGI:1100537; Mid1.
 DR InterPro: IPR003649; Bbox_C.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001870; Gamma_carboxylase.
 DR InterPro: IPR003877; SPRY.
 DR InterPro: IPR003878; SPRY_domain.
 DR InterPro: IPR000315; znf_bbox.
 DR InterPro: IPR001841; znf_ring.
 DR Pfam: PF00041; fn3; 1.
 DR Pfam: PF00622; SPRY; 1.
 DR Pfam: PF00643; zf-B_box; 2.
 DR Pfam: PF00097; zf-C3HC4; 1.
 DR PRINTS: PR01406; BBOXZNFINGER.
 DR SMART: SM00502; BBC; 1.
 DR SMART: SM00336; BBOX; 2.
 DR SMART: SM00060; FN3; 1.
 DR SMART: SM00184; RING; 1.
 DR SMART: SM00449; SPRY; 1.
 DR PROSITE: PS00518; ZINC_FINGER_C3HC4; 1.
 DR Coiled coil; Zinc-finger; Metal-binding; Transcription regulation;
 KW Nuclear protein.
 FT ZN_FING 10 56 C3HC4-TYPE (POTENTIAL).
 FT DOMAIN 170 212 B BOX.
 FT DOMAIN 205 264 COILED COIL (POTENTIAL).
 SQ SEQUENCE 667 AA; 75173 MW; D6EFAICDEA43CBB9 CRC64;
 Query Match 71.4%; Score 35; DB 11; Length 667;
 Best Local Similarity 75.0%; Pred. No. 98;
 Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 KMNHYTVH 8
 DB 444 KQNHVTVH 451
 ID P82458 PRELIMINARY; PRT; 667 AA.
 AC P82458;
 DT 01-JUN-2000 (TrEMBLrel. 14, Created)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE MIDLINE 1 PROTEIN (RING FINGER PROTEIN).
 GN MID1 OR FXV.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99439873; Pubmed=10508587;
 RA Perry J., Ashworth A.;
 RT "Evolutionary rate of a gene affected by chromosomal position.";
 RL Curr. Biol. 9:987-989(1999).
 CC -!- FUNCTION: PUTATIVE TRANSCRIPTIONAL REGULATOR.
 CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
 CC -!- SIMILARITY: CONTAINS A C3HC4-CLASS ZINC FINGER.
 DR EMBL: AF186461; AAD56247.1; -
 DR InterPro: IPR003649; Bbox_C.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR001870; Gamma_carboxylase.
 DR InterPro: IPR003877; SPRY.
 DR InterPro: IPR003878; SPRY_domain.
 DR InterPro: IPR000315; znf_bbox.


```
Query Match 71.4%; Score 35; DB 11; Length 680;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNEYTVH 8
DB 457 KKHVTVH 464

RESULT 15
Q9UJV3
ID Q9UJV3 PRELIMINARY; PRT; 685 AA.
AC Q9UJV3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RING FINGER PROTEIN.
GN FXY2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20112752; PubMed=1064436;
RA Perry J., Short K.M., Romer J.T., Swift S., Cox T.C., Ashworth A.;
RT "FXY2/MID2, a gene related to the X-linked Opitz syndrome gene
RT FXY/MID1, maps to Xq22 and encodes a FNIII domain-containing protein
RT that associates with microtubules.";
RL Genomics 62:385-394(1999).
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL; AF196481; AAF07341.1; -.
DR InterPro; IPR003649; Bbox_C.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003877; SPRY.
DR InterPro; IPR003878; SPRY_domain.
DR InterPro; IPR000315; Znf_bbox.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00622; SPRY; 1.
DR Pfam; PF00643; zf-B_box; 2.
DR SMART; SM00097; zf-C3HC4; 1.
DR SMART; SM00502; BBox; 2.
DR SMART; SM00336; BBOX; 2.
DR SMART; SM00184; RING; 1.
DR SMART; SM00449; SPRY; 1.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; UNKNOWN_1.
KW zinc-finger.
SQ SEQUENCE 685 AA; 77918 MW; 7A68284D23A19EF6 CRC64;

Query Match 71.4%; Score 35; DB 4; Length 685;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNEYTVH 8
DB 444 KKHVTVH 451

RESULT 16
Q9QUS6
ID Q9QUS6 PRELIMINARY; PRT; 685 AA.
AC Q9QUS6;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MIDLINE 2 PROTEIN.
GN MID2 OR FXY2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

Query Match 71.4%; Score 35; DB 11; Length 680;
Best Local Similarity 75.0%; Pred. No. 99;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNEYTVH 8
DB 457 KKHVTVH 464

RESULT 17
Q9UJR9
ID Q9UJR9 PRELIMINARY; PRT; 715 AA.
AC Q9UJR9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MIDLINE 2 PROTEIN.
GN MID2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99330546; PubMed=10400986;
RA Buchner G., Montini E., Andolfi G., Quaderi N., Cainarca S.,
RA Messali S., Bassi M.T., Ballabio A., Meroni G., Franco B.;
RT "MID2, a homologue of the Opitz syndrome gene MID1: similarities in a
RT sub-cellular localization and differences in expression during
RT development.";
RL Hum. Mol. Genet. 8:1397-1407(1999).

Query Match 71.4%; Score 35; DB 11; Length 685;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNEYTVH 8
DB 444 KKHVTVH 451
```

```
NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A.
MEDLINE=99330546; PubMed=10400986;
Buchner G., Montini E., Andolfi G., Quaderi N., Cainarca S.,
Messali S., Bassi M.T., Ballabio A., Meroni G., Franco B.;
"MID2, a homologue of the Opitz syndrome gene MID1: similarities in a
sub-cellular localization and differences in expression during
development.";
Hum. Mol. Genet. 8:1397-1407(1999).
[2]
SEQUENCE FROM N.A.
Perry J., Short K.M., Romer J.T., Swift S., Cox T.C., Ashworth A.;
"Fxy2, a gene related to the X-linked Opitz syndrome gene FXY/MID1,
maps to Xq22 and encodes a FNIII domain-containing protein which
associates with microtubules.";
Genomics 0:0-0(2000).
-!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
EMBL; Y1881; CAB56170.1; -.
DR EMBL; AF196480; AAF07340.1; -.
DR MGD; MGI:1344333; Mid2.
DR InterPro; IPR003649; Bbox_C.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001870; Gamma_carboxylase.
DR InterPro; IPR003877; SPRY.
DR InterPro; IPR003878; SPRY_domain.
DR InterPro; IPR000315; Znf_bbox.
DR InterPro; IPR001841; Znf_ring.
DR Pfam; PF00041; fn3; 1.
DR Pfam; PF00622; SPRY; 1.
DR Pfam; PF00643; zf-B_box; 2.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00502; BBox; 1.
DR SMART; SM00336; BBOX; 2.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00449; SPRY; 1.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; UNKNOWN_1.
KW zinc-finger.
SQ SEQUENCE 685 AA; 77841 MW; 0DA2386C004909A0 CRC64;

Query Match 71.4%; Score 35; DB 11; Length 685;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 KNEYTVH 8
DB 444 KKHVTVH 451

RESULT 17
Q9UJR9
ID Q9UJR9 PRELIMINARY; PRT; 715 AA.
AC Q9UJR9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE MIDLINE 2 PROTEIN.
GN MID2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99330546; PubMed=10400986;
RA Buchner G., Montini E., Andolfi G., Quaderi N., Cainarca S.,
RA Messali S., Bassi M.T., Ballabio A., Meroni G., Franco B.;
RT "MID2, a homologue of the Opitz syndrome gene MID1: similarities in a
RT sub-cellular localization and differences in expression during
RT development.";
RL Hum. Mol. Genet. 8:1397-1407(1999).
```

CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.
DR EMBL; Y18880; CAB56154.1; -.
DR InterPro; IPR003649; BBox.C.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR001870; Gamma_carboxylase.
DR InterPro; IPR003877; SPRY.
DR InterPro; IPR003878; SPRY_domain.
DR InterPro; IPR000315; Znf_box.
DR Pfam; PF00622; SPRY; 1.
DR Pfam; PF00643; zf-B_box; 2.
DR Pfam; PF00097; zf-C3HC4; 1.
DR SMART; SM00502; BBox; 1.
DR SMART; SM00336; BBox; 2.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00184; RING; 1.
DR SMART; SM00449; SPRY; 1.
DR PROSITE; PS00518; ZINC_FINGER_C3HC4; UNKNOWN_1.
KW zinc-finger.
SQ SEQUENCE 715 AA; 81263 MW; B10518806E358BFC CRC64;

Query Match 71.4%; Score 35; DB 4; Length 715;
Best Local Similarity 75.0%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 KMNEYTVH 8
DB 474 KQHHTVH 481
| | | | |

RESULT 18
Q99KL9 PRELIMINARY; PRT; 270 AA.
AC Q99KL9;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE SIMILAR TO REPLICATION PROTEIN A2.
GN RPA2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004578; AAH04578.1; -.
DR MGD; MGI:1339939; Rpa2.
DR InterPro; IPR004365; tRNA_antl.
DR Pfam; PF01336; tRNA_antl; 1.
SQ SEQUENCE 270 AA; 29430 MW; F868B1313CC10DF4 CRC64;

Query Match 69.4%; Score 34; DB 11; Length 270;
Best Local Similarity 62.5%; Pred. No. 62;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 MNEYTVHL 9
DB 152 MNEFTAH 159
| | | | |

RESULT 19
Q92ES2 PRELIMINARY; PRT; 421 AA.
AC Q92ES2;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE LINO386 PROTEIN.
GN LINO386.
OS Listeria innocua.

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
BAQUERO F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
CHARBIT A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
DOMANN E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
ENTIAN K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
GAUTIER L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
JONES L.-M., Kaerst U., Kref J., Kuhn M., Kung F., Kurapkut G.,
MADUENO E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
NORDSIEK G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
REMMELE B., Rose M., Schlueter T., Simoes N., Trierre P.,
RA Vazquez-Boland J.-A., Voss H., Wehlend J., Cossart P.;
RT *Comparative genomics of Listeria species.*;
RL Science 294:849-852(2001).
DR EMBL; AL596164; CAC95619.1; -.
DR Listlist; LINO386; -.
KW Complete proteome.
SQ SEQUENCE 421 AA; 46703 MW; A4F2C723CB0DEA09 CRC64;

Query Match 69.4%; Score 34; DB 16; Length 421;
Best Local Similarity 85.7%; Pred. No. 97;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTV 7
DB 391 KNEYTV 397
| | | | |

RESULT 20
O97291 PRELIMINARY; PRT; 1946 AA.
AC O97291;
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOHETICAL 231.8 KDA PROTEIN.
GN PFC0960C. MAL3P7.21.
OS Plasmodium falciparum (isolate 3D7).
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=36329;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=3D7;
RX MEDLINE=99376085; PubMed=10448855;
RA Bowman S., Lawson D., Basham D., Brown D., Chillingworth T.,
RA Churcher C.M., Craig A., Davies R.M., Devlin K., Feltwell T.,
RA Horrocks P., Jagers K., Jassal B., Kyes S., McLean J., Moulé S.,
RA Mungall K., Murphy L., Oliver K., Quail M.A., Rajandream M.-A.,
RA Rutter S., Skelton J., Squares R., Squares S., Sulston J.E.,
RA Whitehead S., Woodward J.R., Newbold C., Barrell B.G.;
RT "The complete nucleotide sequence of chromosome 3 of Plasmodium falciparum";
RL Nature 400:532-538(1999).
DR EMBL; AL034559; CAB39037.2; -.
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_2.
KW Hypothetical protein.
SQ SEQUENCE 1946 AA; 231792 MW; 59AC248AB9808E34 CRC64;

Query Match 69.4%; Score 34; DB 5; Length 1946;
Best Local Similarity 66.7%; Pred. No. 4.6e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 1181 KMNEYVIFL 1189
| | | | |

```

RESULT 21
Q9KF88      PRELIMINARY;      PRT;      218 AA.
ID Q9KF88;
AC Q9KF88;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BH0597 PROTEIN.
GN BH0597.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001509; BAB04316.1; -.
DR InterPro: IPR003148; KTN
DR InterPro: IPR000309; TrkA_Kuptake.
DR Pfam: PF02254; KTN; 1.
DR Pfam: PF02080; TrkA; 1.
KW Complete proteome.
SQ SEQUENCE 218 AA; 24104 MW; 42F5602106DC9895 CRC64;

Query Match 67.3%; Score 33; DB 16; Length 218;
Best Local Similarity 62.5%; Pred. No. 79;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
Db 37 KVNEFTTH 44
|:|:|:|

RESULT 22
Q9KDN9      PRELIMINARY;      PRT;      227 AA.
ID Q9KDN9;
AC Q9KDN9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TWO-COMPONENT RESPONSE REGULATOR.
GN BH1172.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis."
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -1- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
CC -1- SIMILARITY: TO BACTERIAL REGULATORY PROTEINS INVOLVED IN SIGNAL
CC TRANSDUCTION.
DR EMBL; AF001511; BAB04891.1; -.
DR HSSP; P08402; 1B00.
DR InterPro: IPR001789; Response_reg.
DR InterPro: IPR003662; sub_transporter.

```

```

DR InterPro: IPR001867; Trans_reg_C.
DR Pfam: PF00072; response_reg; 1.
DR Pfam: PF00486; trans_reg_C; 1.
DR SMART; SK00448; REC; 1.
DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
KW Complete proteome; DNA-binding; Phosphorylation; Sensory transduction;
KW Transcription regulation.
SQ SEQUENCE 227 AA; 26094 MW; E9465DC68EF9CB69 CRC64;

Query Match 67.3%; Score 33; DB 16; Length 227;
Best Local Similarity 75.0%; Pred. No. 83;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
Db 1 MSEYLVHL 8
|:|:|:|

RESULT 23
Q96V06      PRELIMINARY;      PRT;      327 AA.
ID Q96V06;
AC Q96V06;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MATING TYPE PROTEIN MAT1-1-1.
GN MAT1-1-1.
OS Cryphonectria parasitica (Chesnut blight fungus) (Endothia
OS parasitica).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Diaporthales; Valsaceae; Cryphonectria.
OX NCBI_TaxID=5116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21546984; PubMed=11686678;
RA McGuire I.C., Marra R.E., Turgeon B.G., Milgroom M.G.;
RT "Analysis of mating-type genes in the chestnut blight fungus,
RT Cryphonectria parasitica."
RL Fungal Genet. Biol. 34:131-144(2001).
DR EMBL; AF380365; AAK83346.1; -.
SQ SEQUENCE 327 AA; 36722 MW; DF0DC3883B1BF4FC CRC64;

Query Match 67.3%; Score 33; DB 3; Length 327;
Best Local Similarity 55.6%; Pred. No. 1-2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db 180 KMNEHQLHI 188
|:|:|:|

RESULT 24
O74547      PRELIMINARY;      PRT;      422 AA.
ID O74547;
AC O74547;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE HYPOTHETICAL 46.9 KDA PROTEIN.
GN SPCC777.08C.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Seeger K., Harris D., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031532; CAA20712.1; -.
KW Hypothetical protein.
SQ SEQUENCE 422 AA; 46863 MW; FDEE90C6C2AE9BE2 CRC64;

```

Query Match 67.3%; Score 33; DB 3; Length 422;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|:|:|:|:|
DB 257 KMNEFVGH 265

RESULT 25

Q9K8I0 ID Q9K8I0 PRELIMINARY; PRT; 525 AA.
AC Q9K8I0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE OLIGOPEPTIDE ABC TRANSPORTER (OLIGOPEPTIDE-BINDING PROTEIN).
GN BH3026.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001517; BAB06745.1; -;
DR InterPro: IPR000914; SBP_bac_5.
DR Pfam: PF00496; SBP_bac_5; 1.
KW Complete proteome.
SQ SEQUENCE 525 AA; 58436 MW; DCB5F3369663F4EF CRC64;

Query Match 67.3%; Score 33; DB 16; Length 525;
Best Local Similarity 62.5%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
:|:|:|:|:
DB 147 IDEYTVHI 154

RESULT 26

Q9WYQ8 ID Q9WYQ8 PRELIMINARY; PRT; 664 AA.
AC Q9WYQ8;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE OXIDOREDUCTASE, PUTATIVE.
GN TM0427.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Geinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).

DR EMBL: AE001721; AAD35512.1; -;
DR TIGR: TM0427; -;
KW Complete proteome.
SQ SEQUENCE 664 AA; 75467 MW; 6CCF1757B87B5CF0 CRC64;

Query Match 67.3%; Score 33; DB 16; Length 664;
Best Local Similarity 55.6%; Pred. No. 2.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|:|:|:|:|
DB 7 KLEYTLRI 15

RESULT 27

Q9A7E2 ID Q9A7E2 PRELIMINARY; PRT; 677 AA.
AC Q9A7E2;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TONB-DEPENDENT RECEPTOR.
GN CCI781.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Niernan W.C., Feldblyum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL: AE005852; AAK23757.1; -;
DR TIGR: CCI781; -;
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
KW Receptor; Complete proteome.
SQ SEQUENCE 677 AA; 73230 MW; 9B54BE1553124C7 CRC64;

Query Match 67.3%; Score 33; DB 16; Length 677;
Best Local Similarity 55.6%; Pred. No. 2.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|:|:|:|:|
DB 280 KMGQYSLHL 288

RESULT 28

Q96CC7 ID Q96CC7 PRELIMINARY; PRT; 776 AA.
AC Q96CC7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SIMILAR TO RAS PROTEIN ACTIVATOR LIKE 1 (GAP1 LIKE).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG CARCINOMA;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

```
DR EMBL; BC014420; AAH14420.1; -.
SQ SSQUENCE 776 AA; 86827 MW; 4BA9067A8D437DF1 CRC64;

Query Match 67.3%; Score 33; DB 4; Length 776;
Best Local Similarity 100.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9
    |||||
Db 56 EYTVHL 61

RESULT 29
Q99K69 PRELIMINARY; PRT; 799 AA.
AC Q99K69;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE RAS PROTEIN ACTIVATOR LIKE 1 (GAP1 LIKE).
GN RASAL1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, GROSS
RC TISSUE.;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005418; AAH05418.1; -.
DR HSSP; P04410; 1A25.
DR MGD; MGI:1330842; Rasal1.
DR InterPro; IPR001562; BTK.
DR InterPro; IPR000008; C2.
DR InterPro; IPR001849; PH.
DR InterPro; IPR001936; RasGAP.
DR Pfam; PF00779; BTK; 1.
DR Pfam; PF00168; C2; 2.
DR Pfam; PF00169; PH; 1.
DR Pfam; PF00616; RasGAP; 1.
DR SMART; SM00107; BTK; 1.
DR SMART; SM00239; C2; 2.
DR SMART; SM00233; PH; 1.
DR SMART; SM00323; RasGAP; 1.
DR PROSITE; PS00499; C2_DOMAIN_1; UNKNOWN_2.
DR PROSITE; PS50004; C2_DOMAIN_2; 2.
DR PROSITE; PS50003; PH_DOMAIN; 1.
DR PROSITE; PS50018; RAS_GTPASE_ACTIV.2; 1.
SQ SEQUENCE 799 AA; 89394 MW; E42239D3B89D8069 CRC64;

Query Match 67.3%; Score 33; DB 11; Length 799;
Best Local Similarity 100.0%; Pred. No. 3e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9
    |||||
Db 56 EYTVHL 61

RESULT 30
Q96129 PRELIMINARY; PRT; 1308 AA.
AC Q96129;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PREDICTED MEMBRANE ASSOCIATED PROTEIN.
GN PFB0125C.
OS Plasmodium falciparum.
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
OX NCBI_TaxID=5833;
```

```

[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=99021743; PubMed=9804551;
RA Gardner M.J., Tettelin H., Carucci D.J., Cummings L.M., Aravind L.,
RA Koonin E.V., Shalloo S., Mason T., Yu K., Fujii C., Pederson J.,
RA Shen K., Jing J., Aston C., Lai Z., Schwartz D.C., Perle M.,
RA Salzberg S., Zhou L., Sutton G.G., Clayton R., White O., Smith H.O.,
RA Fraser C.M., Adams M.D., Venter J.C., Hoffman S.L.;
RT "Chromosome 2 sequence of the human malaria parasite Plasmodium
RT falciparum.";
RL Science 282:1126-1132(1998).
DR EMBL; AE001374; AAC71815.1; -.
DR InterPro; IPR001313; PUM.
SQ SEQUENCE 1308 AA; 155585 MW; 9722F0336606C366 CRC64;

Query Match 67.3%; Score 33; DB 5; Length 1308;
Best Local Similarity 66.7%; Pred. No. 4.9e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
    |||||
Db 1146 KMNEYTKHV 1154

RESULT 31
Q95CP1 PRELIMINARY; PRT; 47 AA.
AC Q95CP1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NATURASE K (FRAGMENT).
GN MATK.
OS Jacksonia horrida.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Mirbelieae; Jacksonia.
OX NCBI_TaxID=140928;
RN [1]
RP SEQUENCE FROM N.A.
RA Chandler G.T., Bayer R.J., Crisp M.D.;
RT "A molecular phylogeny of the endemic Australian genus Gastrolobium
RT (Fabaceae: Mirbelieae) and allied genera using chloroplast and nuclear
RT markers.";
RL Am. J. Bot. 88:1675-1687(2001).
DR EMBL; AF298481; AAL27949.1; -.
KW Chloroplast.
FT NON_TER
SQ SEQUENCE 47 AA; 5736 MW; 574181E4MD93FE08 CRC64;

Query Match 65.3%; Score 32; DB 8; Length 47;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
    |||||
Db 1 MEETVYL 8

RESULT 32
Q99UE1 PRELIMINARY; PRT; 49 AA.
AC Q99UE1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L33.
GN RPMG OR SAS042 OR SAV1335.
OS Staphylococcus aureus (strain N315), and
OS Staphylococcus aureus (strain Mu50).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
```

OX NCBI_TaxID=158879, 158878;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=S.aureus (strain N315), and S.aureus (strain Mu50);
 RX MEDLINE=21311952; PubMed=11418146;
 RA Kuroda M., Ohca T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Lian J.-Q., Ito T.,
 RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hosoyama A.,
 RA Mizutani-Uji Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
 RA Sekimizu K., Hirakawa H., Kuhara S., Goto S., Yabuzaki J.,
 RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
 RA Hattori M., Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 aureus.";
 RL Lancet 357:1225-1240(2001).
 CC -!- SIMILARITY: BELONGS TO THE L33P FAMILY OF RIBOSOMAL PROTEINS.
 DR EMBL; AP003133; BAB2428.1; -;
 DR EMBL; AP003362; BAB57497.1; -;
 DR InterPro; IPR001705; Ribosomal_L33.
 DR Pfam; PF00471; Ribosomal_L33; 1.
 DR ProDom; PD002595; Ribosomal_L33; 1.
 DR PROSITE; PS00582; RIBOSOMAL_L33; 1.
 KW Complete proteome; Ribosomal protein.
 SQ SEQUENCE 49 AA; 5932 MW; DFE4F92CC6AD357D CRC64;

 Query Match 65.3%; Score 32; DB 16; Length 49;
 Best Local Similarity 50.0%; Pred. No. 28;
 Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

 QY 1 KMNEYTVH 8
 DB 38 RLNKYTLH 45
 :::::

RESULT 33
 Q97F09 ID Q97F09 PRELIMINARY; PRT; 130 AA.
 AC Q97F09;
 DT 01-OCT-2001 (TReMBLrel. 18, Created)
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN CAC2946.
 GN CAC2946.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VRM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Berton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf K.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL; AE007791; AAK80888.1; -;
 DR InterPro; IPR003975; SHALCHANNEL.
 DR PRINTS; PR01497; SHALCHANNEL.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 130 AA; 15300 MW; 20E1182F91B18A20 CRC64;

 Query Match 65.3%; Score 32; DB 16; Length 130;
 Best Local Similarity 71.4%; Pred. No. 75;
 Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NEYTVHL 9
 DB 29 NEYEIHL 35
 :::::

RESULT 34
 Q91DR7 ID Q91DR7 PRELIMINARY; PRT; 133 AA.
 AC Q91DR7;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Sapporo virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
 OC Sapporo-like viruses.
 OX NCBI_TaxID=95342;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HU/SIV/CHIBA/990727S/1999;
 RA Okada M., Shinozaki K., Ogawa T., Kaiho I.;
 RT "Molecular epidemiology and phylogenetic analysis of Sapporo-like
 RT viruses in Chiba prefecture, Japan.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ412795; CAC48325.1; -;
 FT NON_TER 1
 FT NON_TER 133
 SQ SEQUENCE 133 AA; 13829 MW; E71FCE4B24026BB9 CRC64;

 Query Match 65.3%; Score 32; DB 12; Length 133;
 Best Local Similarity 55.6%; Pred. No. 76;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

 QY 1 KMNEYTVHL 9
 DB 95 RINPYTAHL 103
 ::|||

RESULT 35
 Q91DP2 ID Q91DP2 PRELIMINARY; PRT; 133 AA.
 AC Q91DP2;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
 DE POLYPROTEIN (FRAGMENT).
 OS Sapporo virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
 OC Sapporo-like viruses.
 OX NCBI_TaxID=95342;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HU/SIV/CHIBA/010469F/2001;
 RA Okada M., Shinozaki K., Ogawa T., Kaiho I.;
 RT "Molecular epidemiology and phylogenetic analysis of Sapporo-like
 RT viruses in Chiba prefecture, Japan.";
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ412820; CAC48350.1; -;
 FT NON_TER 1
 FT NON_TER 133
 SQ SEQUENCE 133 AA; 13875 MW; FADF86B301D8BB9 CRC64;

 Query Match 65.3%; Score 32; DB 12; Length 133;
 Best Local Similarity 55.6%; Pred. No. 76;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
 DB 95 RINPYTAHL 103
 ::|||

RESULT 36
 Q91DP1 ID Q91DP1 PRELIMINARY; PRT; 133 AA.
 AC Q91DP1;
 DT 01-DEC-2001 (TReMBLrel. 19, Created)
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)


```
DE POLYPROTEIN (FRAGMENT).
OS Sapporo virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Sapporo-like viruses.
OX NCBI_TaxID=95342;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HU/SLV/CHIBA/010557S/2001;
RA Okada M., Shinozaki K., Ogawa T., Kaiho I.;
RT "Molecular epidemiology and phylogenetic analysis of Sapporo-like
RT viruses in Chiba prefecture, Japan.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ412821; CAC48351.1; -.
FT NON_TER 1
FT NON_TER 133
SQ SEQUENCE 133 AA; 13996 MW; 9744B9D38674CEC3 CRC64;

Query Match 65.3%; Score 32; DB 12; Length 133;
Best Local Similarity 55.6%; Pred. No. 76;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db 95 RINPYTAHL 103

RESULT 37
Q91DN8
ID Q91DN8 PRELIMINARY; PRT; 133 AA.
AC Q91DN8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Sapporo virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Sapporo-like viruses.
OX NCBI_TaxID=95342;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HU/SLV/CHIBA/010592F/2001;
RA Okada M., Shinozaki K., Ogawa T., Kaiho I.;
RT "Molecular epidemiology and phylogenetic analysis of Sapporo-like
RT viruses in Chiba prefecture, Japan.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ412824; CAC48354.1; -.
FT NON_TER 1
FT NON_TER 133
SQ SEQUENCE 133 AA; 13996 MW; 9744B9D38674CEC3 CRC64;

Query Match 65.3%; Score 32; DB 12; Length 133;
Best Local Similarity 55.6%; Pred. No. 76;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db 95 RINPYTAHL 103

RESULT 38
Q91DN6
ID Q91DN6 PRELIMINARY; PRT; 133 AA.
AC Q91DN6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Sapporo virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Sapporo-like viruses.
OX NCBI_TaxID=95342;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=HU/SLV/CHIBA/010604F/2001;
RA Okada M., Shinozaki K., Ogawa T., Kaiho I.;
RT "Molecular epidemiology and phylogenetic analysis of Sapporo-like
RT viruses in Chiba prefecture, Japan.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ412826; CAC48356.1; -.
FT NON_TER 1
FT NON_TER 133
SQ SEQUENCE 133 AA; 13996 MW; 9744B9D38674CEC3 CRC64;

Query Match 65.3%; Score 32; DB 12; Length 133;
Best Local Similarity 55.6%; Pred. No. 76;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db 95 RINPYTAHL 103

RESULT 39
Q91DR6
ID Q91DR6 PRELIMINARY; PRT; 134 AA.
AC Q91DR6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Sapporo virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Sapporo-like viruses.
OX NCBI_TaxID=95342;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HU/SLV/CHIBA/990763S/1999;
RA Okada M., Shinozaki K., Ogawa T., Kaiho I.;
RT "Molecular epidemiology and phylogenetic analysis of Sapporo-like
RT viruses in Chiba prefecture, Japan.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ412796; CAC48326.1; -.
FT NON_TER 1
FT NON_TER 134
SQ SEQUENCE 134 AA; 13890 MW; 99B492DF96FED691 CRC64;

Query Match 65.3%; Score 32; DB 12; Length 134;
Best Local Similarity 55.6%; Pred. No. 77;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
Db 96 RINPYTAHL 104

RESULT 40
Q91DR5
ID Q91DR5 PRELIMINARY; PRT; 134 AA.
AC Q91DR5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Sapporo virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Sapporo-like viruses.
OX NCBI_TaxID=95342;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HU/SLV/CHIBA/991172S/1999;
RA Okada M., Shinozaki K., Ogawa T., Kaiho I.;
RT "Molecular epidemiology and phylogenetic analysis of Sapporo-like
RT viruses in Chiba prefecture, Japan.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ412797; CAC48327.1; -.
FT NON_TER 1
```

```
FT NON_TER 134 134
SQ SEQUENCE 134 AA; 13905 MW; B4B492DF82780DE8 CRC64;

Query Match
Best Local Similarity 65.3%; Score 32; DB 12; Length 134;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KMNEYTVHL 9
Db 96 RINPYTAHL 104

RESULT 41
Q91DP6 PRELIMINARY; PRT; 134 AA.
AC Q91DP6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Sapporo virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Sapporo-like viruses.
OX NCBI_TaxID=95342;
RN [1]
RC STRAIN=HU/SLV/CHIBA/010004H/2001;
RA Okada M., Shinozaki K., Ogawa T., Kaiho I.;
RT "Molecular epidemiology and phylogenetic analysis of Sapporo-like
viruses in Chiba prefecture, Japan.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ412816; CAC48346.1; -.
FT NON_TER 1 134
FT NON_TER 134 134
SQ SEQUENCE 134 AA; 13905 MW; B4B492DF82780DE8 CRC64;

Query Match
Best Local Similarity 65.3%; Score 32; DB 12; Length 134;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KMNEYTVHL 9
Db 96 RINPYTAHL 104

RESULT 42
Q91DN4 PRELIMINARY; PRT; 134 AA.
AC Q91DN4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE POLYPROTEIN (FRAGMENT).
OS Sapporo virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Caliciviridae;
OC Sapporo-like viruses.
OX NCBI_TaxID=95342;
RN [1]
RC STRAIN=HU/SLV/CHIBA/010675F/2001;
RA Okada M., Shinozaki K., Ogawa T., Kaiho I.;
RT "Molecular epidemiology and phylogenetic analysis of Sapporo-like
viruses in Chiba prefecture, Japan.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ412828; CAC48358.1; -.
FT NON_TER 1 134
FT NON_TER 134 134
SQ SEQUENCE 134 AA; 13905 MW; B4B492DF82780DE8 CRC64;

Query Match
Best Local Similarity 65.3%; Score 32; DB 12; Length 134;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KMNEYTVHL 9
Db 96 RINPYTAHL 104

RESULT 43
O57714 PRELIMINARY; PRT; 144 AA.
ID O57714
AC O57714;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 16.5 KDA PROTEIN PH1990.
GN PH1990.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RC STRAIN=OT3;
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000007; BAA31117.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 144 AA; 16467 MW; 148827603DCD053F CRC64;

Query Match
Best Local Similarity 71.4%; Score 32; DB 17; Length 144;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KMNEYTV 7
Db 47 KMNEYSI 53

RESULT 44
Q9GZ14 PRELIMINARY; PRT; 181 AA.
ID Q9GZ14
AC Q9GZ14;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PES-1 (FRAGMENT).
GN PES-1.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RC STRAIN=FROM N.A.
RX MEDLINE=20500977; PubMed=11044397;
RA Molin L., Mounsey A., Aslam S., Bauer P., Young J., James M.,
RA Sharma-Oates A., Hope I.A.;
RT "Evolutionary conservation of redundancy between a diverged pair of
for-head transcription factor homologues.";
RL Development 127:4825-4835(2000).
DR EMBL; AF260299; AAF99288.1; -.
DR HSSP; Q63245; 2HFH.
DR InterPro; IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
FT NON_TER 181 181
```

```
OY 1 KMNEYTVHL 9
Db 96 RINPYTAHL 104

RESULT 43
O57714 PRELIMINARY; PRT; 144 AA.
ID O57714
AC O57714;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 16.5 KDA PROTEIN PH1990.
GN PH1990.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RC STRAIN=FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL; AP000007; BAA31117.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 144 AA; 16467 MW; 148827603DCD053F CRC64;

Query Match
Best Local Similarity 71.4%; Score 32; DB 17; Length 144;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

OY 1 KMNEYTV 7
Db 47 KMNEYSI 53

RESULT 44
Q9GZ14 PRELIMINARY; PRT; 181 AA.
ID Q9GZ14
AC Q9GZ14;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PES-1 (FRAGMENT).
GN PES-1.
OS Caenorhabditis briggsae.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6238;
RN [1]
RC STRAIN=FROM N.A.
RX MEDLINE=20500977; PubMed=11044397;
RA Molin L., Mounsey A., Aslam S., Bauer P., Young J., James M.,
RA Sharma-Oates A., Hope I.A.;
RT "Evolutionary conservation of redundancy between a diverged pair of
for-head transcription factor homologues.";
RL Development 127:4825-4835(2000).
DR EMBL; AF260299; AAF99288.1; -.
DR HSSP; Q63245; 2HFH.
DR InterPro; IPR001766; Fork_head.
DR Pfam; PF00250; Fork_head; 1.
DR PRINTS; PR00053; FORKHEAD.
DR SMART; SM00339; FH; 1.
DR PROSITE; PS00658; FORK_HEAD_2; 1.
DR PROSITE; PS50039; FORK_HEAD_3; 1.
FT NON_TER 181 181
```

```
SQ SEQUENCE 181 AA; 19157 MW; B6BCACBDF0861BFC CRC64;
Query Match 65.3%; Score 32; DB 5; Length 181;
Best Local Similarity 71.4%; Pred. NO. 1e+02;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTV 7
Db 37 KMNEFTI 43
|||||:

RESULT 45
Q9PMC4
ID Q9PMC4 PRELIMINARY; PRT; 192 AA.
AC Q9PMC4;
DT 01-OCT-2000 (TREMELrel. 15, Created)
DT 01-OCT-2000 (TREMELrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)
DE MDAB PROTEIN HOMOLOG.
GN CJ1545C.
OS Campylobacter jejuni.
OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
OC Campylobacter.
OX NCBI_TaxID=197;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCTC 11168;
RX MEDLINE=20150912; PubMed=10688204;
RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
RA Whitehead S., Barrell B.G.;
RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
RT reveals hypervariable sequences.";
RL Nature 403:665-668(2000).
DR EMBL; AL139078; CAB73961.1; -.
DR InterPro; IPR003680; NADHdh_2.
DR Pfam; PF02525; NADHdh_2; 1.
KW Complete proteome.
SQ SEQUENCE 192 AA; 22223 MW; 5DC47E175A60FB36 CRC64;

Query Match 65.3%; Score 32; DB 16; Length 192;
Best Local Similarity 62.5%; Pred. NO. 1.1e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
Db 181 LNEYELHL 188
|||||:
```

Search completed: November 6, 2002, 12:11:54
Job time : 21.4444 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 11 Seconds
(without alignments)
78.619 Million cell updates/sec

Title: US-09-905-083-31
Perfect score: 49
Sequence: 1 KMNEYTVHL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	49	100.0	253	2 A53968	serine proteinase
2	39	79.6	218	1 YVBPk3	lysis protein t -
3	39	79.6	218	1 YVBPt4	lysis protein t -
4	37	75.5	334	2 B98019	conserved hypotet
5	37	75.5	393	2 D75207	hypothetical prote
6	36	73.5	627	2 T25395	hypothetical prote
7	36	73.5	1829	2 T34239	hypothetical prote
8	35	71.4	159	2 A11406	spermidine/spermin
9	35	71.4	159	2 A11782	spermidine/spermin
10	35	71.4	505	2 H75431	conserved hypotet
11	35	71.4	555	2 AD1794	acylase and dieste
12	35	71.4	667	2 T09482	ring finger protei
13	35	71.4	667	2 T09013	RING finger protei
14	34	69.4	270	2 A43711	replication protei
15	34	69.4	270	2 S28682	replication protei
16	34	69.4	285	2 F70348	hypothetical prote
17	34	69.4	421	2 AC1481	conserved hypotet
18	34	69.4	421	2 AH1120	B. subtilis ywbn p
19	33	67.3	204	2 B31227	nitrogen fixation
20	33	67.3	204	2 E95345	FixJ transcription
21	33	67.3	218	2 S53354	calflagin Tb-24 -
22	33	67.3	218	2 E83724	hypothetical prote
23	33	67.3	227	2 D83796	two-component resp
24	33	67.3	229	2 S53355	calflagin Tb-1.7 -
25	33	67.3	233	1 AQUT17	flagellar calcium-
26	33	67.3	407	2 S53353	calflagin Tb-44A -
27	33	67.3	422	2 T11714	hypothetical prote
28	33	67.3	429	2 H70307	preprotein translo
29	33	67.3	525	2 B84028	oligopeptide ABC t

30	33	67.3	664	2 C72379	hypothetical prote
31	33	67.3	677	2 A87470	TonB-dependent rec
32	33	67.3	1042	2 S23738	pyr1-3 protein - s
33	33	67.3	1308	2 E71622	probable membrane
34	33	67.3	1481	1 QZDOP3	pyrimidine synthe
35	32	65.3	49	2 H89908	50S ribosomal prot
36	32	65.3	84	2 AE2172	hypothetical prote
37	32	65.3	116	2 S24989	ribosomal protein
38	32	65.3	130	2 E97262	hypothetical prote
39	32	65.3	144	2 F71215	hypothetical prote
40	32	65.3	192	2 G81301	MdaB protein homol
41	32	65.3	228	2 C90033	hypothetical prote
42	32	65.3	304	2 H75378	hypothetical prote
43	32	65.3	339	2 F97121	probable membrane
44	32	65.3	371	2 A89800	conserved hypotet
45	32	65.3	448	2 F82280	citrate/sodium sym

ALIGNMENTS

RESULT 1
A53968

serine proteinase SCCE precursor - human

N:Alternate names: stratum corneum chymotryptic enzyme

C:Species: Homo sapiens (man)

C>Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999

C:Accession: A53968

R: Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, J. Biol. Chem. 269, 19420-19426, 1994

A: Title: Cloning, expression, and characterization of stratum corneum chymotryptic en

A: Reference number: A53968; MUID: 94308225

A: Accession: A53968

A: Status: preliminary

A: Molecule type: mRNA

A: Residues: 1-253 <HAN>

A: Cross-references: GB:L33404; NID: g521214; PIDN: AAC37551.1; PID: g532504

C: Genetics:

A: Gene: GDB: PRSS6; SCCE

A: Cross-references: GDB: 377730

A: Map position: 7q35-7q35

C: Superfamily: trypsin; trypsin homology

F: 30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 49; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.024; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 KMNEYTVHL 9
|||||
Db 72 KMNEYTVHL 80

RESULT 2

YVBPk3

lysis protein t - phage K3

C: Species: phage K3

A: Note: host Escherichia coli

C: Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 23-Jul-1999

C: Accession: A27083

R: Riede, I.

J. Bacteriol. 169, 2956-2961, 1987

A: Title: Lysis gene t of T-even bacteriophages: evidence that colicins and bacterioph

A: Reference number: A27083; MUID: 87250254

A: Accession: A27083

A: Molecule type: DNA

A: Residues: 1-218 <RIE>

A: Cross-references: GB:M16812; NID: g215503; PIDN: AAA88415.1; PID: g215504

A: Note: the author translated the codon CAA for residue 85 as Ile and CAG for residue

C: Genetics:

A: Gene: t

C: Superfamily: phage T4 lysis protein t

C: Keywords: host cell lysis

Query Match 79.6%; Score 39; DB 1; Length 218;
Best Local Similarity 87.5%; Pred. No. 2.2;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
|:|||||
Db 139 MDEYTVHL 146

RESULT 3

YVBT4

Lysis protein t - phase T4

C:Species: phage T4

A:Note: host Escherichia coli

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 23-Jul-1999

C:Accession: JF0028; S07395

R:Montag, D.; Degen, M.; Henning, U.

Nucleic Acids Res. 15, 6736, 1987

A:Title: Nucleotide sequence of gene t (lysis gene) of the E. coli phage T4.

A:Reference number: S07395; MUID:87316934

A:Accession: JF0028

A:Molecule type: DNA

A:Residues: 1-218 <NON>

A:Cross-references: GB:Y00408; NID:g15368; PIDN:CAA68470.1; PID:g15369

A:Note: the sequence is almost identical with that of the E. coli phage K3

C:Comment: At the end of the growth cycle, phage T4 expresses two genes with lysis function about the gene product of t, although it has been suggested that it acts as a phospholipase.

C:Genetics:

A:Gene: t

A:Map position: 157,985-158,639

C:Superfamily: phage T4 lysis protein t

C:Keywords: host cell lysis; transmembrane protein

F:35-49/Domain: transmembrane #status predicted <TMN>

Query Match 79.6%; Score 39; DB 1; Length 218;
Best Local Similarity 87.5%; Pred. No. 2.2;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
|:|||||
Db 139 MDEYTVHL 146

RESULT 4

B98019

conserved hypothetical protein sprl179 [imported] - Streptococcus pneumoniae (strain R6)

C:Species: Streptococcus pneumoniae

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001

C:Accession: B98019

R:Hoskins, J.A.; Alborn Jr., W.; Arnold, J.; Blaszcak, L.; Burgett, S.; DeHoff, B.S.; E

e, R.; LeBlanc, D.J.; Lee, L.N.; Lefkowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M

y, P.; Sun, P.M.; Winkler, M.B.

J. Bacteriol. 183, 5709-5717, 2001

A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;

A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.

A:Reference number: A97872; MUID:21429245; PMID:11544234

A:Accession: B98019

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-334 <KUR>

A:Cross-references: GB:AE007317; PIDN:AAK99982.1; PID:g15458811; GSPDB:GN00174

C:Genetics:

A:Gene: sprl179

Query Match 75.5%; Score 37; DB 2; Length 334;
Best Local Similarity 66.7%; Pred. No. 8.9;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|:|||||
Db 302 KMNEYITHI 310

RESULT 5

D75207

hypothetical protein PAB2235 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 28-Jul-2000

C:Accession: D75207

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome s

A:Reference number: A75001

A:Accession: D75207

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-393 <KAW>

A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49107.1; PID:g545

A:Experimental source: strain Orsay

C:Genetics:

A:Gene: PAB2235

C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0196

Query Match 75.5%; Score 37; DB 2; Length 393;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|:|||||
Db 129 KLPEYTHL 137

RESULT 6

T25395

hypothetical protein T28A8.6 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 20-Jun-2000

C:Accession: T25395

R:Lloyd, C.

submitted to the EMBL Data Library, March 1997

A:Reference number: Z20027

A:Accession: T25395

A>Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

A:Residues: 1-627 <WIL>

A:Cross-references: EMBL:Z92813; PIDN:CAB07289.1; GSPDB:GN00021; CESP:T28A8.6

A:Experimental source: clone T28A8

C:Genetics:

A:Gene: CESP:T28A8.6

A:Map position: 3

A:Introns: 51/2; 89/2; 183/2; 221/2; 236/2; 607/3

C:Superfamily: Caenorhabditis elegans hypothetical protein Y57A10A.d

Query Match 73.5%; Score 36; DB 2; Length 627;
Best Local Similarity 66.7%; Pred. No. 28;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|:|||||
Db 62 KMNEYSIEL 70

RESULT 7

T34239

hypothetical protein F26F12.7 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999

C:Accession: T34239

R:Wilson, R.; Bentley, D.; Gattung, S.

submitted to the EMBL Data Library, April 1996

A:Description: The sequence of C. elegans cosmid F26F12.

A:Reference number: Z21493

A:Accession: T34239

A>Status: preliminary; translated from GB/EMBL/DBBJ

A:Molecule type: DNA

T09482

ring finger protein FXY - human

C:Species: Homo sapiens (man)

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 02-Sep-2000

C:Accession: T09482

R:Perry, J.; Feather, S.; Smith, A.; Palmer, S.; Ashworth, A.

submitted to the EMBL Data Library, November 1997

A:Description: The human FXY maps to chromosome Xp22.3: Implications for evolution of th

A:Reference number: Z16687

A:Accession: T09482

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-667 <PER>

A:Cross-references: EMBL:AF035360; NID:g2827993; PID:g2827994

C:Genetics:

A:Gene: FXY

A:Map position: xp22.3

C:Superfamily: RING finger homology

F;6-65/Domain: RING finger homology <RRN>

Query Match 71.4%; Score 35; DB 2; Length 667;

Best Local Similarity 75.0%; Pred. No. 48; Mismatches 2; Indels 0; Gaps 0;

Matches 6; Conservative 0;

QY 1 KMNEYTVH 8

| | | | |

Db 444 QNHVTVH 451

RESULT 13

T09013

RING finger protein Fxy - mouse

C:Species: Mus musculus (house mouse)

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 02-Sep-2000

C:Accession: T09013

R:Palmer, S.; Perry, J.; Kipling, D.; Ashworth, A.

Proc. Natl. Acad. Sci. U.S.A. 94, 12030-12035, 1997

A:Title: A gene spans the pseudautosomal boundary. in mice.

A:Reference number: Z16531; MUID:98004518

A:Accession: T09013

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: mRNA

A:Residues: 1-667 <PAL>

A:Cross-references: EMBL:AF026565; NID:g2589222; PIDN:AAB83986.1; PID:g2589223

C:Genetics:

A:Gene: Fxy

A:Map position: x; y

C:Superfamily: RING finger homology

C:Keywords: zinc finger

F;6-65/Domain: RING finger homology <RRN>

Query Match 71.4%; Score 35; DB 2; Length 667;

Best Local Similarity 75.0%; Pred. No. 48; Mismatches 2; Indels 0; Gaps 0;

Matches 6; Conservative 0;

QY 1 KMNEYTVH 8

| | | | |

Db 444 QNHVTVH 451

RESULT 14

A43711

replication protein repA 32K chain - human

C:Species: Homo sapiens (man)

C>Date: 03-Mar-1993 #sequence_revision 03-Mar-1993 #text_change 16-Feb-2001

C:Accession: A43711

R:Erdlie, L.F.; Wold, M.S.; Kelly, T.J.

J. Biol. Chem. 265, 3177-3182, 1990

A:Title: The primary structure of the 32-kDa subunit of human replication protein A.

A:Reference number: A43711; MUID:90153966

A:Accession: A43711

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-270 <ERD>

A:Cross-references: EMBL:J05249; NID:g337349; PIDN:AAA36560.1; PID:g337350

C:Superfamily: Schizosaccharomyces pombe single-stranded DNA binding protein 30K chai

Query Match 69.4%; Score 34; DB 2; Length 270;

Best Local Similarity 62.5%; Pred. No. 29; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 2;

QY 2 MNEYTVHL 9

| | | | |

Db 152 MNEFTTHI 159

RESULT 15

S28682

replication protein A 32K chain homolog - mouse

C:Species: Mus musculus (house mouse)

C>Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 16-Feb-2001

C:Accession: S28682

R:Nakagawa, M.; Tsukada, S.; Soma, T.; Shimizu, Y.; Miyake, S.; Iwamatsu, A.; Sugiyam

Nucleic Acids Res. 19, 4292, 1991

A:Title: cDNA cloning of the murine 30-kDa protein homologous to the 32-kDa subunit o

A:Reference number: S28682; MUID:91334146

A:Accession: S28682

A:Molecule type: mRNA

A:Residues: 1-270 <NAG>

A:Cross-references: EMBL:D00812; NID:g220583; PIDN:BAA00693.1; PID:g220584

C:Superfamily: Schizosaccharomyces pombe single-stranded DNA binding protein 30K chai

Query Match 69.4%; Score 34; DB 2; Length 270;

Best Local Similarity 62.5%; Pred. No. 29; Mismatches 1; Indels 0; Gaps 0;

Matches 5; Conservative 2;

QY 2 MNEYTVHL 9

| | | | |

Db 152 MNEFTTHI 159

RESULT 16

F70348

hypothetical protein aq_539 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999

C:Accession: F70348

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.;

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: F70348

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-285 <AQF>

A:Cross-references: GB:AE000694; NID:g2983162; PIDN:AAC06778.1; PID:g2983177; GB:AE00

A:Experimental source: strain VFS

C:Genetics:

A:Gene: aq_539

Query Match 69.4%; Score 34; DB 2; Length 285;

Best Local Similarity 66.7%; Pred. No. 31; Mismatches 2; Indels 0; Gaps 0;

Matches 6; Conservative 2;

QY 1 KMNEYTVHL 9

| | | | |

Db 207 KFNEYTLNL 215

RESULT 17

AC1481

conserved hypothetical protein B. subtilis YWBN protein homolog lin0386 [imported] -

C:Species: Listeria innocua

C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C:Accession: AC1481

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AC1481
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC95619.1; PID:gl6412815; GSPDB:GN00178
A:Experimental source: strain Clp11262
C:Genetics:
A:Gene: lin0386
C:Superfamily: hypothetical protein ycdB

Query Match 69.4%; Score 34; DB 2; Length 421;
Best Local Similarity 85.7%; Pred. No. 47;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTV 7
Db 391 KNEYTV 397
|:|||||

RESULT 18
AH1120
B. subtilis ywb protein homolog lmo0367 [imported] - *Listeria* monocytogenes (strain EGI
C:Species: *Listeria* monocytogenes
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001
C:Accession: AH1120
R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Karst, U.
Science 294, 849-852, 2001
A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of *Listeria* species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AH1120
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-421 <GLA>
A:Cross-references: GB:NC_003210; PIDN:CAC98446.1; PID:gl6409745; GSPDB:GN00177
A:Experimental source: strain EGD-e
C:Genetics:
A:Gene: lmo0367
C:Superfamily: hypothetical protein ycdB

Query Match 69.4%; Score 34; DB 2; Length 421;
Best Local Similarity 85.7%; Pred. No. 47;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTV 7
Db 391 KNEYTV 397
|:|||||

RESULT 19
B31227
nitrogen fixation regulatory protein fixJ - *Rhizobium* meliloti
C:Species: *Rhizobium* meliloti
C:Date: 28-Feb-1990 #sequence_revision 28-Feb-1990 #text_change 29-Sep-1999
C:Accession: B31227; S32838; S39985
R:David, M.; Davenport, M.L.; Batut, J.; Dedieu, A.; Domergue, O.; Chai, J.; Hertig, C.; H
Cell 54, 671-683, 1988
A:Title: Cascade regulation of nif gene expression in *Rhizobium* meliloti.
A:Reference number: A90901; MUID:88311069
A:Accession: B31227
A:Molecule type: DNA
A:Residues: 1-204 <DAV>

A:Cross-references: EMBL:221854; NID:g49403; PIDN:CAA79898.1; PID:g49405
C:Genetics:
A:Gene: fixJ
C:Superfamily: nitrogen fixation regulatory protein fixJ; response regulator homology
C:Keywords: DNA binding; phosphoprotein; transcription regulation
F:6-115/Domain: response regulator homology <RRH>
F:54/Binding site: phosphate (Asp) (covalent) #status predicted

Query Match 67.3%; Score 33; DB 2; Length 204;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
Db 1 MTDYTVHI 8
|:|||||

RESULT 20
E95345
FixJ Transcriptional activator [imported] - *Sinorhizobium meliloti* (strain 1021) mag
C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: E95345
R:Barnett, M.J.; Fisher, R.F.; Jones, T.; Komp, C.; Abola, A.P.; Barloy-Hubler, F.; B
.; Kalman, S.; Keating, D.H.; Palm, C.; Peck, M.C.; Surzycki, R.; Wells, D.H.; Yeh, K.
Proc. Natl. Acad. Sci. U.S.A. 98, 9883-9888, 2001
A:Title: Nucleotide sequence and predicted functions of the entire *Sinorhizobium meli*
A:Reference number: A95262; MUID:21396509; PMID:11481432
A:Accession: E95345
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <KUR>
A:Cross-references: GB:AE006469; PIDN:AAK65327.1; PID:gl4523784; GSPDB:GN00165
A:Experimental source: strain 1021, megaplasmid pSymA
R:Galibert, F.; Finan, T.M.; Long, S.R.; Punler, A.; Abola, P.; Ampe, F.; Barloy-Hub
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: fixJ
A:Genome: plasmid
C:Superfamily: nitrogen fixation regulatory protein fixJ; response regulator homology

Query Match 67.3%; Score 33; DB 2; Length 204;
Best Local Similarity 62.5%; Pred. No. 34;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
Db 1 MTDYTVHI 8
|:|||||

RESULT 21
S53354
calflagin Tb-24 - *Trypanosoma brucei*
N:Alternate names: EF-hand calcium-binding protein
C:Species: *Trypanosoma brucei*
C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 13-Aug-1999
C:Accession: S53354
R:Wu, Y.; Deford, J.; Benjamin, R.; Lee, M.G.S.; Ruben, L.
Biochem. J. 304, 833-841, 1994
A:Title: The gene family of EF-hand calcium-binding proteins from the flagellum of *T*
A:Reference number: S53353; MUID:95118301
A:Accession: S53354
A:Molecule type: DNA
A:Residues: 1-218 <WUY>
A:Cross-references: EMBL:U06644; NID:g458438; PIDN:AAB40004.1; PID:g458439
C:Superfamily: flagellar calcium-binding protein; calmodulin repeat homology

C:Keywords: calcium binding; EF hand
F:167-199/Domain: calmodulin repeat homology <EF4>

Query Match 67.3%; Score 33; DB 2; Length 218;
Best Local Similarity 55.6%; Pred. No. 37;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
|::|:| |
Db 82 KLDEFTHL 90

RESULT 22

E83724

hypothetical protein BH0597 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: E83724

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: E83724

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-218 <STO>

A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04316.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0597

C:Superfamily: conserved hypothetical protein MG323

Query Match 67.3%; Score 33; DB 2; Length 218;

Best Local Similarity 62.5%; Pred. No. 37;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8

|::|:| |

Db 37 KVNEFTTH 44

RESULT 23

D83796

two-component response regulator BH1172 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: D83796

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: D83796

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-227 <STO>

A:Cross-references: GB:AP001511; GB:BA000004; NID:g10173727; PIDN:BA04891.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH1172

C:Superfamily: ompR protein; response regulator homology

Query Match 67.3%; Score 33; DB 2; Length 227;

Best Local Similarity 75.0%; Pred. No. 38;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9

|::|:| |

Db 1 MSEVLVHL 8

RESULT 24

S53355

calflagin Tb-1.7 - Trypanosoma brucei (fragment)

N:Alternate names: EF-hand calcium-binding protein
C:Species: Trypanosoma brucei

C:Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 13-Aug-1999

C:Accession: S53355

R:Wu, Y.; Deford, J.; Benjamin, R.; Lee, M.G.S.; Ruben, L.

Biochem. J. 304, 833-841, 1994

A:Title: The gene family of EF-hand calcium-binding proteins from the flagellum of Tr

A:Reference number: S53353; MUID:95118301

A:Accession: S53355

A:Molecule type: DNA

A:Residues: 1-229 <WUY>

A:Cross-references: EMBL:U05882; NID:g453390; PIDN:AAA75582.1; PID:g453391

C:Superfamily: flagellar calcium-binding protein; calmodulin repeat homology

C:Keywords: calcium binding; EF hand

F:163-195/Domain: calmodulin repeat homology <EF4>

Query Match 67.3%; Score 33; DB 2; Length 229;

Best Local Similarity 55.6%; Pred. No. 39;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

|::|:| |

Db 78 KLDEFTHL 86

RESULT 25

AQU17

flagellar calcium-binding protein Tb17 - Trypanosoma brucei

C:Species: Trypanosoma brucei

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 22-Jun-1999

C:Accession: S10515

R:Lee, M.G.S.; Chen, J.; Ho, A.W.M.; d'Alesandro, P.A.; van der Ploeg, L.H.T.

Nucleic Acids Res. 18, 4252, 1990

A:Title: A putative flagellar Ca²⁺-binding protein of the flagellum of trypanosomatid

A:Reference number: S10515; MUID:90332427

A:Accession: S10515

A:Molecule type: mRNA

A:Residues: 1-233 <LEE>

A:Cross-references: EMBL:X53464; NID:g10540; PIDN:CAA37558.1; PID:g10541

A>Note: the authors translated the codon TCA for residue 4 as Ala, CGG for residue 30-

6 as Arg

C:Superfamily: flagellar calcium-binding protein; calmodulin repeat homology

C:Keywords: calcium binding; duplication; EF hand; flagellum

F:48-80/Domain: calmodulin repeat homology <EF1>

F:97-129/Domain: calmodulin repeat homology <EF2>

F:130-162/Domain: calmodulin repeat homology <EF3>

F:167-199/Domain: calmodulin repeat homology <EF4>

Query Match 67.3%; Score 33; DB 1; Length 233;

Best Local Similarity 55.6%; Pred. No. 39;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

|::|:| |

Db 82 KLDEFTHL 90

RESULT 26

S53353

calflagin Tb-44A - Trypanosoma brucei

N:Alternate names: EF-hand calcium-binding protein

C:Species: Trypanosoma brucei

C:Date: 28-Oct-1996 #sequence_revision 07-Feb-1997 #text_change 13-Aug-1999

C:Accession: S53353; S71064

R:Wu, Y.; Deford, J.; Benjamin, R.; Lee, M.G.S.; Ruben, L.

Biochem. J. 304, 833-841, 1994

A:Title: The gene family of EF-hand calcium-binding proteins from the flagellum of Tr

A:Reference number: S53353; MUID:95118301

A:Accession: S53353

A:Molecule type: mRNA

A:Residues: 1-407 <WUY>

A:Cross-references: EMBL:U06463

A>Note: the authors translated the codon CTT for residue 263 as J and AAG for residue

R:Deford, J.H.

submitted to the EMBL Data Library, February 1994

A:Reference number: S71064

A:Accession: S71064

A:Molecule type: mRNA

A:Residues: 1-161, 'W', 163-350, 'W', 352-407 <DEF>

A:Cross-references: EMBL:U06463; NID:g995575; PIDN:AAA75583.1; PID:g458430

C:Superfamily: flagellar calcium-binding protein; calmodulin repeat homology

C:Keywords: calcium binding; EF hand

F:48-80/Domain: calmodulin repeat homology <EF1>

F:97-129/Domain: calmodulin repeat homology <EF2>

F:130-162/Domain: calmodulin repeat homology <EF3>

F:167-199/Domain: calmodulin repeat homology <EF4>

F:237-269/Domain: calmodulin repeat homology <EF5>

F:286-318/Domain: calmodulin repeat homology <EF6>

F:319-351/Domain: calmodulin repeat homology <EF7>

F:356-388/Domain: calmodulin repeat homology <EF8>

Query Match 67.3%; Score 33; DB 2; Length 407;

Best Local Similarity 55.6%; Pred. No. 72;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

|||||

Db 271 KLDEFTHL 279

RESULT 27

Tl1714

hypothetical protein SPCC777.08c - fission yeast (*Schizosaccharomyces pombe*)

C:Species: *Schizosaccharomyces pombe*

C>Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 16-Jul-1999

C:Accession: Tl1714

R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.

submitted to the EMBL Data Library, August 1998

A:Reference number: Tl7318

A:Accession: Tl1714

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-422 <SEE>

A:Cross-references: EMBL:AL031532; NID:el319424; PID:el319432

A:Experimental source: strain 972h(-)

C:Genetics:

A:Map position: IIIR

A:Note: SPCC777.08c

Query Match 67.3%; Score 33; DB 2; Length 422;

Best Local Similarity 66.7%; Pred. No. 75;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

|||||

Db 257 KMNEFVGH 265

RESULT 28

H70307

preprotein translocase SecY - *Aquifex aeolicus*

C:Species: *Aquifex aeolicus*

C>Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999

C:Accession: H70307

V:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*.

A:Reference number: A70300; MUID:98196666

A:Accession: H70307

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-429 <AQF>

A:Cross-references: GB:AE000672; NID:g2982810; PIDN:AAC06435.1; PID:g2982812; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

A:Gene: secY

C:Superfamily: preprotein translocase secY

Query Match 67.3%; Score 33; DB 2; Length 429;

Best Local Similarity 66.7%; Pred. No. 76;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

|||||

Db 111 KINEYTKYL 119

RESULT 29

B84028

oligopeptide ABC transporter (oligopeptide-binding protein) BH3026 [imported] - *Bacill*

C:Species: *Bacillus halodurans*

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 22-Oct-2001

C:Accession: B84028

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B84028

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-525 <STO>

A:Cross-references: GB:AP001517; GB:BA000004; NID:g10175500; PIDN:BAB06745.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

C:Superfamily: dipeptide transport protein

Query Match 67.3%; Score 33; DB 2; Length 525;

Best Local Similarity 62.5%; Pred. No. 95;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9

|||||

Db 147 IDEYTVHI 154

RESULT 30

C72379

hypothetical protein TM0427 - *Thermotoga maritima* (strain MSB8)

C:Species: *Thermotoga maritima*

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000

C:Accession: C72379

R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic

Garrett, M.M.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson, M.

C.M.

Nature 399, 323-329, 1999

A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome

A:Reference number: A72200; MUID:99287316

A:Accession: C72379

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-664 <ARN>

A:Cross-references: GB:AE001721; GB:AE000512; NID:g4980922; PIDN:AAD35512.1; PID:g498

A:Experimental source: strain MSB8

C:Genetics:

A:Gene: TM0427

Query Match 67.3%; Score 33; DB 2; Length 664;

Best Local Similarity 55.6%; Pred. No. 1.2e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9

|||||

Db 7 KLNEYTLRI 15

RESULT 31

AB7470

TonB-dependent receptor [imported] - Caulobacter crescentus

C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: A87470
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: A87470
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-677 <STO>
A:Cross-references: GB:AE005673; NID:g13423209; PIDN:AAK23757.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1781

Query Match 67.3%; Score 33; DB 2; Length 677;
Best Local Similarity 55.6%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 280 KMGQYSLHL 288
|| |::||
RESULT 32
S23738
PY1-3 protein - slime mold (Dictyostellium discoideum) (fragment)
N:Contains: carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5)
C:Species: Dictyostellium discoideum
C:Date: 08-Jun-1994 #sequence_revision 08-Sep-1995 #text_change 09-Jun-2000
C:Accession: S23738
R:Elgar, G.; Schofield, J.P.
DNA Seq. 2, 219-226, 1992
A:Title: Carbamoyl phosphate synthetase (CPSase) in the PY1-3 multigene of Dictyostellium
A:Reference number: S23738; MUID:92329976
A:Accession: S23738
A:Molecule type: DNA
A:Residues: 1-1042 <ELG>
A:Cross-references: EMBL:X55433; NID:g7244; PIDN:CAA39077.1; PID:g7245
C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology; Ba
arbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-phos
C:Keywords: ligase
F:1-1042/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homolo
F:1-1042/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
F:1-433/Domain: biotin carboxylase homology <BC1>
F:543-990/Domain: biotin carboxylase homology <BC2>

Query Match 67.3%; Score 33; DB 2; Length 677;
Best Local Similarity 55.6%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 280 KMGQYSLHL 288
|| |::||
RESULT 32
S23738
PY1-3 protein - slime mold (Dictyostellium discoideum) (fragment)
N:Contains: carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5)
C:Species: Dictyostellium discoideum
C:Date: 08-Jun-1994 #sequence_revision 08-Sep-1995 #text_change 09-Jun-2000
C:Accession: S23738
R:Elgar, G.; Schofield, J.P.
DNA Seq. 2, 219-226, 1992
A:Title: Carbamoyl phosphate synthetase (CPSase) in the PY1-3 multigene of Dictyostellium
A:Reference number: S23738; MUID:92329976
A:Accession: S23738
A:Molecule type: DNA
A:Residues: 1-1042 <ELG>
A:Cross-references: EMBL:X55433; NID:g7244; PIDN:CAA39077.1; PID:g7245
C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology; Ba
arbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-phos
C:Keywords: ligase
F:1-1042/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homolo
F:1-1042/Domain: carbamoyl-phosphate synthase (ammonia) homology <CPA>
F:1-433/Domain: biotin carboxylase homology <BC1>
F:543-990/Domain: biotin carboxylase homology <BC2>

Query Match 67.3%; Score 33; DB 2; Length 1042;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 988 KMTENTIHL 996
|| |::||
RESULT 33
E71622
probable membrane associated protein PFB0125c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71622
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pterea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: E71622
A:Status: preliminary
A:Molecule type: DNA

Query Match 67.3%; Score 33; DB 2; Length 1308;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 1146 KNEYTKHV 1154
|| |::||
RESULT 34
QZDOP3
pyrimidine synthesis protein PY1-3 - slime mold (Dictyostellium discoideum) (fragment)
N:Contains: aspartate carbamoyltransferase (EC 2.1.3.2); carbamoyl-phosphate synthase
C:Species: Dictyostellium discoideum
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 07-Aug-1998
C:Accession: S02800
R:Faure, M.; Camonis, J.H.; Jacquet, M.
Eur. J. Biochem. 179, 345-356, 1989
A:Title: Molecular characterization of a Dictyostellium discoideum gene encoding a mul
A:Reference number: S02800; MUID:89137111
A:Accession: S02800
A:Molecule type: DNA
A:Residues: 1-467;468-1481 <FAU>
A:Cross-references: EMBL:X14633
A:Experimental source: strain AX3
C:Genetics:
A:Gene: PY1-3
C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology;
arbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-p
C:Keywords: hydrolase; ligase; methyltransferase; multifunctional enzyme; pyrimidine
F:1-707/Domain: carbamoyl-phosphate synthase (ammonia) homology (fragments) <CPA>
F:1-340/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homo
F:158-340/Domain: trpG homology <TRG>
F:357-704/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain ho
F:367-467/Domain: biotin carboxylase homology (fragment) <BC1>
F:468-651/Domain: biotin carboxylase homology (fragment) <BC2>
F:721-1066/Domain: Bacillus dihydroorotase homology <DHO>
F:1179-1477/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
F:236/Active site: Cys #status predicted

A:Residues: 1-1308 <GAR>
A:Cross-references: GB:AE001374; GB:AE001362; NID:g3845100; PIDN:AACT71815.1; PID:g384
A:Experimental source: clone 3D7
C:Genetics:
A:Gene: PFB0125c

Query Match 67.3%; Score 33; DB 2; Length 1308;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 1146 KNEYTKHV 1154
|| |::||
RESULT 34
QZDOP3
pyrimidine synthesis protein PY1-3 - slime mold (Dictyostellium discoideum) (fragment)
N:Contains: aspartate carbamoyltransferase (EC 2.1.3.2); carbamoyl-phosphate synthase
C:Species: Dictyostellium discoideum
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 07-Aug-1998
C:Accession: S02800
R:Faure, M.; Camonis, J.H.; Jacquet, M.
Eur. J. Biochem. 179, 345-356, 1989
A:Title: Molecular characterization of a Dictyostellium discoideum gene encoding a mul
A:Reference number: S02800; MUID:89137111
A:Accession: S02800
A:Molecule type: DNA
A:Residues: 1-467;468-1481 <FAU>
A:Cross-references: EMBL:X14633
A:Experimental source: strain AX3
C:Genetics:
A:Gene: PY1-3
C:Superfamily: rudimentary enzyme; aspartate/ornithine carbamoyltransferase homology;
arbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain homology; carbamoyl-p
C:Keywords: hydrolase; ligase; methyltransferase; multifunctional enzyme; pyrimidine
F:1-707/Domain: carbamoyl-phosphate synthase (ammonia) homology (fragments) <CPA>
F:1-340/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) small chain homo
F:158-340/Domain: trpG homology <TRG>
F:357-704/Domain: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain ho
F:367-467/Domain: biotin carboxylase homology (fragment) <BC1>
F:468-651/Domain: biotin carboxylase homology (fragment) <BC2>
F:721-1066/Domain: Bacillus dihydroorotase homology <DHO>
F:1179-1477/Domain: aspartate/ornithine carbamoyltransferase homology <ACT>
F:236/Active site: Cys #status predicted

Query Match 67.3%; Score 33; DB 1; Length 1481;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 649 KMTENTIHL 657
|| |::||
RESULT 35
H89908
50S ribosomal protein L33 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: H89908
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; O
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, R.; Inoue, R.; Kaito, C.; Sekimizu, K.
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: H89908
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-49 <KUR>
A:Cross-references: GB:BA000018; PID:g13701133; PIDN:BAB42428.1; GSPDB:GN00149
A:Experimental source: strain N315

Query Match 67.3%; Score 33; DB 2; Length 1042;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
DB 988 KMTENTIHL 996
|| |::||
RESULT 33
E71622
probable membrane associated protein PFB0125c - malaria parasite (Plasmodium falciparum)
C:Species: Plasmodium falciparum
C:Date: 13-Nov-1998 #sequence_revision 13-Nov-1998 #text_change 21-Jul-2000
C:Accession: E71622
R:Gardner, M.J.; Tettelin, H.; Carucci, D.J.; Cummings, L.M.; Aravind, L.; Koonin, E.V.;
; Pterea, M.; Salzberg, S.; Zhou, L.; Sutton, G.G.; Clayton, R.; White, O.; Smith, H.O.
Science 282, 1126-1132, 1998
A:Title: Chromosome 2 sequence of the human malaria parasite Plasmodium falciparum.
A:Reference number: A71600; MUID:99021743
A:Accession: E71622
A:Status: preliminary
A:Molecule type: DNA

Query Match 67.3%; Score 33; DB 2; Length 1042;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

C;Genetics:
A;Gene: rpmG

Query Match 65.3%; Score 32; DB 2; Length 49;
Best Local Similarity 50.0%; Pred. No. 12;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
:::|::|::|

Db 38 RLNKYTLH 45

RESULT 36

AE2172

hypothetical protein asr2932 [imported] - Anabaena sp. (strain PCC 7120)

C;Species: Anabaena sp.

A;Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002

C;Accession: AE2172

R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana

A;Reference number: AB1807; MUID:21595285; PMID:11759840

A;Accession: AE2172

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-84 <KUR>

A;Cross-references: GB:BA000019; PIDN:BA874631.1; PID:gl17132026; GSPDB:GN00179

A;Experimental source: strain PCC 7120

C;Genetics:

A;Gene: asr2932

Query Match

Best Local Similarity 65.3%; Score 32; DB 2; Length 84;

C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NEYTVHL 9

|||::|::|

Db 6 NEYAVHM 12

RESULT 37

S24989

ribosomal protein L31.e, cytosolic - Chlamydomonas reinhardtii

C;Species: Chlamydomonas reinhardtii

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 13-Aug-1999

C;Accession: S24989

R;Dumont, F.

submitted to the EMBL Data Library, May 1992

A;Reference number: S24989

A;Accession: S24989

A;Molecule type: mRNA

A;Residues: 1-116 <DUM>

A;Cross-references: EMBL:X66413; NID:gl8208; PIDN:CAAA7044.1; PID:gl8209

C;Superfamily: rat ribosomal protein L31

C;Keywords: protein biosynthesis; ribosome

Query Match

Best Local Similarity 65.3%; Score 32; DB 2; Length 116;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 4 EYTVHL 9

|||::|::|

Db 15 EYTIHL 20

RESULT 38

E97262

hypothetical protein CAC2946 [imported] - Clostridium acetobutylicum

C;Species: Clostridium acetobutylicum

C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C;Accession: E97262

R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; J

J. Bacteriol. 183, 4823-4838, 2001

A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A;Reference number: A96900; MUID:21359325; PMID:21359325

A;Accession: E97262

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-130 <KUR>

A;Cross-references: GB:AE001437; PIDN:AAK80888.1; PID:gl5025998; GSPDB:GN00168

A;Experimental source: Clostridium acetobutylicum ATCC824

C;Genetics:

A;Gene: CAC2946

Query Match

Best Local Similarity 65.3%; Score 32; DB 2; Length 130;

Matches 5; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 3 NEYTVHL 9

|||::|::|

Db 29 NEYEIHL 35

RESULT 39

F71215

hypothetical protein PH1990 - Pyrococcus horikoshii

C;Species: Pyrococcus horikoshii

C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C;Accession: F71215

R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; S-

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogi

DNA Res. 5, 55-76, 1998

A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A;Reference number: A71000; MUID:98344137

A;Accession: F71215

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-144 <KAW>

A;Cross-references: GB:AP000007; NID:g3236134; PIDN:BAA31117.1; PID:g3258434

A;Experimental source: strain OT3

A;Note: this accession replaces an interim accession for a sequence replaced by GenB

C;Genetics:

A;Gene: PH1990

Query Match

Best Local Similarity 65.3%; Score 32; DB 2; Length 144;

Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 KMNEYTV 7

|||::|::|

Db 47 KMNEYSI 53

RESULT 40

G81301

MdaB protein homolog Cjl1545c [imported] - Campylobacter jejuni (strain NCTC 11168)

C;Species: Campylobacter jejuni

C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 02-Sep-2000

C;Accession: G81301

R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chl

C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVliet, A.; Whitehead, S.; B-

Nature 403, 665-668, 2000

A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals

A;Reference number: A81250; MUID:20150912

A;Accession: G81301

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-192 <PAR>

A;Cross-references: GB:AL139078; GB:AL111168; NID:g968723; PIDN:CAB73961.1; PID:g96

A;Experimental source: serotype O2, strain NCTC 11168

C;Genetics:

A;Gene: Cjl1545c

C;Superfamily: NAD(P)H dehydrogenase (quinone) 2

```

Query Match          65.3%; Score 32; DB 2; Length 192;
Best Local Similarity 62.5%; Pred. No. 51;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
    :|||:|
Db 181 LNEYELHL 188

RESULT 41
C90033
hypothetical protein SA2127 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: C90033
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogasawara, M.; Mizutani-Ul, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.; Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: C90033
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-228 <KUR>
A:Cross-references: GB:BA000018; PID:g13702136; PIDN:BAB43428.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
C:Superfamily: Haemophilus influenzae ribose-5-phosphate isomerase

Query Match          65.3%; Score 32; DB 2; Length 228;
Best Local Similarity 55.6%; Pred. No. 61;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVHL 9
    :|||:|
Db 191 KPHEYLHL 199

RESULT 42
H75378
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
C:Accession: H75378
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.; M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; M.S.; Smith, H.O.; Venter, J.C.; Fraser, C.M. Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: H75378
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-304 <WHI>
A:Cross-references: GB:AE002002; GB:AE000513; NID:g6459345; PIDN:RAF11154.1; PID:g645935
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1591
A:Map position: 1
C:Superfamily: Deinococcus radiodurans hypothetical protein DR1591

Query Match          65.3%; Score 32; DB 2; Length 304;
Best Local Similarity 75.0%; Pred. No. 84;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 MNEYTVHL 9
    :|||:|
Db 289 LNEYTVSL 296
```

```

RESULT 43
F97121
Probable membrane-associated Zn-dependent protease [imported] - Clostridium acetobuty-
C:Species: Clostridium acetobutylicum
C>Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: F97121
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R. J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: F97121
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-339 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79761.1; PID:g15024768; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC1796

Query Match          65.3%; Score 32; DB 2; Length 339;
Best Local Similarity 62.5%; Pred. No. 94;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
    :|||:|
Db 145 KINYSVH 152

RESULT 44
A89800
conserved hypothetical protein SA0332 [imported] - Staphylococcus aureus (strain N315)
C:Species: Staphylococcus aureus
C>Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C:Accession: A89800
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogasawara, M.; Mizutani-Ul, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.; Lancet 357, 1225-1240, 2001
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A:Reference number: A89758; MUID:21311952; PMID:11418146
A:Accession: A89800
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-371 <KUR>
A:Cross-references: GB:BA000018; PID:g13700258; PIDN:BAB41556.1; GSPDB:GN00149
A:Experimental source: strain N315
C:Genetics:
C:Superfamily: hypothetical protein ycdB

Query Match          65.3%; Score 32; DB 2; Length 371;
Best Local Similarity 62.5%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 KMNEYTVH 8
    :|||:|
Db 340 KNEYIYTH 347

RESULT 45
F82280
citrate/sodium symporter VC0795 [imported] - Vibrio cholerae (strain N16961 serogroup
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82280
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.; chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: F82280
```

```

A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-448 <HEI>
A:Cross-references: GB:AE004164; GB:AE003852; NID:9655232; PIDN:AAF93959.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype E1 Tor
C:Genetics:
A:Gene: VC0795
A:Map position: 1
C:Superfamily: citrate transport protein citC

Query Match      65.3%; Score 32; DB 2; Length 448;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY  2 MNEYTVHL 9
    ||| |::|
DB  1 MNEKTLHL 8

Search completed: November 6, 2002, 12:08:06
Job time : 14 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 18.4444 Seconds
(without alignments)
84.413 Million cell updates/sec

Title: US-09-905-083-32
Perfect score: 40
Sequence: 1 RLSSMKVKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	33	82.5	282	Q9S2Q8	Q9S2q8 streptomyce
2	32	80.0	351	Q97Z26	Q97z26 sulfolobus
3	32	80.0	664	Q9P7W7	Q9p7w7 schizosacch
4	31	77.5	97	Q9B021	Q9b021 bacterioiopa
5	31	77.5	912	Q9LDK6	Q9ldk6 arabidopsis
6	30	75.0	105	Q93WB3	Q93wb3 arabidopsis
7	30	75.0	236	Q93XZ4	Q93xz4 arabidopsis
8	30	75.0	239	Q80845	Q80845 arabidopsis
9	30	75.0	240	Q17203	Q17203 caenorhabdi
10	30	75.0	321	Q70341	Q70341 mus musculus
11	30	75.0	373	Q9V014	Q9v014 pyrococcus
12	30	75.0	391	Q9C6U9	Q9c6u9 arabidopsis
13	30	75.0	394	Q9C857	Q9c857 arabidopsis
14	30	75.0	517	Q94C74	Q94c74 arabidopsis
15	30	75.0	578	Q9VA29	Q9va29 drosophila
16	30	75.0	647	Q49468	Q49468 mycoplasma

17	30	75.0	815	3	Q9HEF3	Q9hef3 neurospora
18	30	75.0	986	17	Q9UXB1	Q9uxb1 sulfolobus
19	30	75.0	1167	16	Q9PAG0	Q9pag0 xylella fas
20	30	75.0	1768	5	Q9N8K7	Q9n8k7 trypanosoma
21	30	75.0	1887	5	Q9GZG6	Q9gzg6 caenorhabdi
22	29	72.5	118	10	Q40388	Q40388 nicotiana b
23	29	72.5	242	10	Q9FG20	Q9fg20 arabidopsis
24	29	72.5	285	16	Q97F67	Q97f67 clostridium
25	29	72.5	336	17	Q26946	Q26946 methanother
26	29	72.5	345	17	Q97ZT5	Q97zt5 sulfolobus
27	29	72.5	347	17	Q976Z8	Q976z8 sulfolobus
28	29	72.5	347	17	Q972T3	Q972t3 sulfolobus
29	29	72.5	347	17	Q972N5	Q972n5 sulfolobus
30	29	72.5	349	17	Q9HR20	Q9hr20 halobacteri
31	29	72.5	357	16	Q97L60	Q97l60 clostridium
32	29	72.5	359	8	Q9TJR6	Q9tjr6 prototheca
33	29	72.5	379	16	Q9RSQ4	Q9rsq4 deinococcus
34	29	72.5	434	17	Q97410	Q97410 sulfolobus
35	29	72.5	434	17	Q972H0	Q972h0 sulfolobus
36	29	72.5	434	17	Q96ZC6	Q96zc6 sulfolobus
37	29	72.5	436	17	Q96XG3	Q96xg3 sulfolobus
38	29	72.5	469	17	Q9HIJ8	Q9hij8 thermoplasm
39	29	72.5	553	16	Q9X1Y5	Q9x1y5 thermotoga
40	29	72.5	578	16	Q9PPR7	Q9ppr7 ureaplasma
41	29	72.5	593	5	Q965D0	Q965d0 theileria p
42	29	72.5	656	5	Q9XU08	Q9xuu8 caenorhabdi
43	29	72.5	691	16	Q07430	Q07430 mycobacteri
44	29	72.5	849	16	Q92S82	Q92s82 rhizobium m
45	29	72.5	1165	5	Q95ZA8	Q95za8 leishmania

ALIGNMENTS

RESULT 1

Q9S2Q8 PRELIMINARY; PRT; 282 AA.
AC Q9S2Q8;
DT 01-MAY-2000 (TREMELrel. 13, Created)
DT 01-MAY-2000 (TREMELrel. 13, Last sequence update)
DE 01-MAY-2000 (TREMELrel. 13, Last annotation update)
DE PUTATIVE ENDO ALPHA-1,4 POLYGALACTOSAMINIDASE.
GN SC5F7.23C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K., Harris D.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Redenbach M., Kisser H.M., Denapaite D., Eichner A., Cullum J., Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL: AL096872; CAB51262.1;
SQ SEQUENCE 282 AA; 30789 MW; BEECFE1743703A33 CRC64;

Query Match 82.5%; Score 33; DB 2; Length 282;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
|||:||||
Db 265 RLSSMLKK 272

RESULT 2

Q97226 PRELIMINARY; PRT; 351 AA.
AC Q97226; (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DE 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE BIOTIN SYNTHASE (BIOB) (EC 2.8.1.6).
GN BIOB.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.,
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006729; AAK41371.1; -.
KW Transferrase; Complete proteome.
SQ SEQUENCE 351 AA; 39814 MW; 892DB90BC987F752 CRC64;

Query Match 80.0%; Score 32; DB 17; Length 351;

Best Local Similarity 66.7%; Pred. No. 58;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 9
|||:||||
Db 341 RLDSLIKK 349

RESULT 3

Q9P7W7 PRELIMINARY; PRT; 664 AA.
AC Q9P7W7; (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL ARMADILLO/BETA-CATENIN DOMAIN PROTEIN.
GN SPBC1703.03C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA McDougall R.C., Rajandream M.A., Barrell B.G., Cadieu E., Lelaure V.,
RA Galibert F.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL136536; CAB66447.1; -.
DR InterPro; IPR000225; Armadillo.
DR Pfam; PF00514; Armadillo_seg; 2.
DR PROSITE; PS50176; ARM_REPEAT; 1.
SQ SEQUENCE 664 AA; 74811 MW; 363FEE00EB15E69F CRC64;

Query Match 80.0%; Score 32; DB 3; Length 664;

Best Local Similarity 66.7%; Pred. No. 1.le+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 9
|||:||||
Db 624 RLKNMVKKI 632

RESULT 4

Q9B021 PRELIMINARY; PRT; 97 AA.
AC Q9B021; (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DE HYPOTHETICAL 11.0 KDA PROTEIN (FRAGMENT).
OS Bacteriophage GMSE-1.
OC Viruses.
OX NCBI_TaxID=148339;
RN [1]
RP SEQUENCE FROM N.A.
RA Dale C., Young S.A., Maudlin I., Welburn S.C.;
RT "Endosymbiont bacteriophage may influence susceptibility to
RT trypanosome infection in tsetse.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF311657; AAG30264.1; -.
KW Hypothetical protein.
FT NON_TER 1
SQ SEQUENCE 97 AA; 10987 MW; B17CB63BE43F2C52 CRC64;

Query Match 77.5%; Score 31; DB 9; Length 97;

Best Local Similarity 87.5%; Pred. No. 27;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LSSMVKK 9
|||:||||
Db 88 LQSMVKK 95

RESULT 5

Q9LDK6 PRELIMINARY; PRT; 912 AA.
AC Q9LDK6; (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-OCT-2001 (TRENBLrel. 18, Last annotation update)
DE HYPOTHETICAL 105.3 KDA PROTEIN.
GN AT4G39750.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Monfort A., Casacuberta E., Puigdomenech P., Mewes H.W., Lemcke K.,
RA Mayer K.F.X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Bevan M., Monfort A., Casacuberta E., Puigdomenech P., Hohelsel J.,
RA Mewes H.W., Lemcke K., Mayer K.F.X.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL161595; CAB80638.1; -.
DR EMBL; AL022605; CAB77062.1; -.
DR InterPro; IPR001810; F-box.
DR InterPro; IPR001798; Kelch.
DR Pfam; PF00646; F-box; 3.
DR Pfam; PF01344; Kelch; 4.
DR SMART; SM00256; FBOX; 2.
DR PROSITE; PS50181; FBOX; 1.
KW Hypothetical protein.

SQ SEQUENCE 912 AA; 105278 MW; 8C764AFDD114D57E CRC64;

Query Match 77.5%; Score 31; DB 10; Length 912;
 Best Local Similarity 55.6%; Pred. No. 2.5e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 | | | | |
 Db 403 RFSMIRKI 411

RESULT 6

Q93W83 ID Q93W83 PRELIMINARY; PRT; 105 AA.

AC Q93W83;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DE AT2645740/F4118.28.

OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Kim C.J., Chen H., Cheuk R., Koesema E., Meyers M.C., Banh J.,
 RA Bowser L., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
 RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
 RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
 RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis ORF clones";
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.

RN [2]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Shinn P.,
 RA Tracy S.E., Banh J., Bowser L., Carninci P., Chung M.K., Kamiya A.,
 RA Goldsmith A.D., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
 RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
 RA Miranda M., Narusaka M., Nguyen M., Palm C.J., Pham P.K., Quach H.L.,
 RA Sakano H., Sakurai T., Satou M., Seki M., Southwick A., Tang C.C.,
 RA Toriumi M., Yamada K., Yu G., Shinozaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AY057732; AAL15362.1; -;
 DR EMBL; AY037208; AAK59793.1; -;
 SQ SEQUENCE 105 AA; 11433 MW; 45F8B7A34096CF4C CRC64;

Query Match 75.0%; Score 30; DB 10; Length 105;
 Best Local Similarity 66.7%; Pred. No. 49;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 | | | | |
 Db 17 RLSSSMKKI 25

RESULT 7

Q93X24 ID Q93X24 PRELIMINARY; PRT; 236 AA.

AC Q93X24;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DE HYPOTHETICAL 25.9 KDA PROTEIN.
 GN AT2645740 OR F4118.28.

OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
 RA Ecker J., Theologis A., Davis R.W.;
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AY054520; AAK96711.1; -;
 KW Hypothetical protein.

SQ SEQUENCE 236 AA; 25943 MW; 102FAE6A197A5378 CRC64;

Query Match 75.0%; Score 30; DB 10; Length 236;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 | | | | |
 Db 148 RLSSSMKKI 156

RESULT 8

O80845 ID O80845 PRELIMINARY; PRT; 239 AA.

AC O80845;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DE AT2645740 PROTEIN.
 GN AT2645740.

OS Arabidopsis thaliana (Mouse-ear cross).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxID=3702;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shear T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen W., VanAken S.E., Umayam L., Fallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana";
 RL Nature 402:761-768(1999).

RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AC004665; AAC28551.1; -;
 SQ SEQUENCE 239 AA; 26221 MW; FABB9C1FED9EDC5B CRC64;

Query Match 75.0%; Score 30; DB 10; Length 239;
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 | | | | |
 Db 148 RLSSSMKKI 156

RESULT 9

O17203 ID O17203 PRELIMINARY; PRT; 240 AA.

AC O17203;

DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 28.4 KDA PROTEIN.
GN C08E3.12.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Miller N., Kramer J., Keppler D.;
RT "The sequence of C. elegans cosmid C08E3.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF025457; AAB70970.1; -;
DR InterPro; IPR002900; DUF38.
DR Pfam; PF01827; DUF38; 1.
KW Hypothetical protein.
SQ SEQUENCE 240 AA; 28372 MW; 6284E8F1195D78E6 CRC64;

Query Match 75.0%; Score 30; DB 5; Length 240;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
I: |::|||
DB 193 RIPSMLKKV 201

RESULT 10
ID O70341 PRELIMINARY; PRT; 321 AA.
AC O70341;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TAPOXIN-ASSOCIATED CALCIUM BINDING PROTEIN 49.
GN RCN2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95239201; PubMed=7722520;
RA Dodds D., Schlimgen A.K., Lu S.Y., Perin M.S.;
RT "Novel reticular calcium binding protein is purified on taipoxin
RT columns.";
RL J. Neurochem. 64:2339-2344(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Perin M.S.;
RT "Mouse taipoxin-associated calcium binding protein 49 (TCBP49).";
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
DR EMBL; AF049125; AAC05132.1; -;
DR MGD; MGI:1349765; Rcn2.
DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000886; ER_target.
DR Pfam; PF00036; ehand; 4.
DR SMART; SM00054; EFh; 3.
DR PROSITE; PS00018; EF_HAND; UNKNOWN_4.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW Calcium-binding.
SQ SEQUENCE 321 AA; 37455 MW; E83323E9B437A7A8 CRC64;

Query Match 75.0%; Score 30; DB 11; Length 321;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
I: |::|||
DB 68 RLQSIKKI 76

RESULT 11
ID Q9V014 PRELIMINARY; PRT; 373 AA.
AC Q9V014;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 43.5 KDA PROTEIN.
GN PAB1719.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248286; CAB49892.1; -;
DR InterPro; IPR000051; SAM_bind.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 373 AA; 43535 MW; E043A552546BAP8 CRC64;

Query Match 75.0%; Score 30; DB 17; Length 373;
Best Local Similarity 77.8%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
I: |::|||
DB 102 RLGSYKKV 110

RESULT 12
ID Q9C6U9 PRELIMINARY; PRT; 391 AA.
AC Q9C6U9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LIPASE, POTATIVE.
GN F27M3_22.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway T.H., Devar K.,
RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,

RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin x., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana";
 RL Nature 408:816-820(2000).
 DR EMBL: AC074360; AAG60153.1; -;
 DR InterPro: IPR001087; Lipase_GDSL.
 SQ SEQUENCE 391 AA; 43001 MW; EE0F05F7FC7B9FDF CRC64;
 Query Match 75.0%; Score 30; DB 10; Length 391;
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LSSMVK 8
 Db 373 LSSMIKK 379
 |||||:
 RESULT 13
 Q9C857 PRELIMINARY; PRT; 394 AA.
 AC Q9C857;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE HYPOTHETICAL 43.4 KDA PROTEIN.
 GN T8E3.19.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Theologis A., Ecker J.R., Falm C.J., Pederspiel N.A., Kaul S.,
 RX STRAIN=CV. COLUMBIA;
 RC MEDLINE=21016719; PubMed=11130712;
 RA Theologis A., Ecker J.R., Falm C.J., Pederspiel N.A., Kaul S.,
 RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
 RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
 RA Chung M.K., Conn L., Conway A.B., Conway A.R., Cressy T.H., Dewar K.,
 RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,
 RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
 RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
 RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
 RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
 RA Lin x., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
 RA Militscher J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
 RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
 RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
 RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
 RA Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
 RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
 RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
 thaliana";
 RL Nature 408:816-820(2000).
 DR EMBL: AC027135; AAG51269.1; -;
 DR InterPro: IPR001087; Lipase_GDSL.
 KW Hypothetical protein.
 SQ SEQUENCE 394 AA; 43375 MW; D90EE1914D4740CD CRC64;
 Query Match 75.0%; Score 30; DB 10; Length 394;
 Best Local Similarity 85.7%; Pred. No. 1.8e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVK 8
 Db 376 LSSMIKK 382
 |||||:
 RESULT 14
 Q94C74 PRELIMINARY; PRT; 517 AA.
 ID Q94C74;
 AC Q94C74;
 DT 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE GLYCINE HYDROXYMETHYLTRANSFERASE.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 OC NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yamada K., Liu S.X., Sakano H., Pham P.K., Banh J., Chung M.K.,
 RA Goldsmith A.D., Lee J.M., Quach H.L., Tang C., Toriumi M., Yu G.,
 RA Bowser L., Carninci P., Chen H., Cheuk R., Hayashizaki Y., Ishida J.,
 RA Jones T., Kamiya A., Karlin-Neumann G., Kawai J., Kim C., Kossema E.,
 RA Lam B., Lin J., Meyers M.C., Miranda M., Narusaka M., Nguyen M.,
 RA Palm C.J., Sakurai T., Satou M., Seki M., Shinn P., Southwick A.,
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "Full length cDNA of gene F2P16_40/AY5926780";
 RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY035117; AAK59622.1; -;
 KW Transferase; Methyltransferase.
 SQ SEQUENCE 517 AA; 57322 MW; 612A1D9703894E21 CRC64;
 Query Match 75.0%; Score 30; DB 10; Length 517;
 Best Local Similarity 87.5%; Pred. No. 2.4e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RLSSMVK 8
 Db 7 RLSSSVKK 14
 |||||:
 RESULT 15
 Q9VA29 PRELIMINARY; PRT; 578 AA.
 ID Q9VA29;
 AC Q9VA29;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE CG12071 PROTEIN.
 GN CG12071.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Eurygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OC NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpén G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskaas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL; AE003775; AAF57095.1; -.
 DR HSP; P07248; 2ADR.
 DR FlyBase; FBgn0039808; CG12071.
 DR InterPro; IPR000822; Znf-C2H2.
 DR Pfam; PF00096; Zf-C2H2; 3.
 DR SMART; SM00355; Znf-C2H2; 3.
 DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 3.
 DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; 3.
 KW DNA-binding; Metal-binding; Zinc-finger.
 SQ SEQUENCE 578 AA; 64636 MW; 19966A0D7BFA43FE CRC64;

Query Match 75.0%; Score 30; DB 5; Length 578;
 Best Local Similarity 75.0%; Pred. No. 2.6e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
 Db 249 RLNNMVKK 256
 ||:|||||

RESULT 16
 Q49468
 ID Q49468 PRELIMINARY; PRT; 647 AA.
 AC Q49468; Q53303;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE HEMAGGLUTININ HOMOLOG PRECURSOR.
 GN PMGA1.2.
 OS Mycoplasma gallisepticum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
 OC Mycoplasmataceae; Mycoplasma.
 OX NCBI_TaxID=2096;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93162830; PubMed=8432610;
 RA Markham P.F., Glew M.D., Whithear K.G., Walker I.D.;
 RT "Molecular cloning of a member of the gene family that encodes pmGA, a
 RT hemagglutinin of *Mycoplasma gallisepticum*.";
 RL Infect. Immun. 61:903-909(1993).
 DR EMBL; M83178; AAA02996.1; -.
 DR EMBL; S55216; AAB25397.2; -.
 KW Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 647 HEMAGGLUTININ HOMOLOG.
 SQ SEQUENCE 647 AA; 70333 MW; 33916673BB9E28C4 CRC64;

Query Match 75.0%; Score 30; DB 2; Length 647;

Best Local Similarity 75.0%; Pred. No. 2.9e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LSSMVKKV 9
 Db 237 LSSIVKKI 244
 |||:|||||
 RESULT 17
 Q9HEF3
 ID Q9HEF3 PRELIMINARY; PRT; 815 AA.
 AC Q9HEF3;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CONSERVED HYPOTHETICAL PROTEIN.
 GN 65E11.130.
 OS Neurospora crassa.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 OC Sordariales; Sordariaceae; Neurospota.
 OX NCBI_TaxID=5141;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Schulte U., Aign V., Hoheisel J., Brandt P., Fartmann B., Holland R.,
 RA Nyakatura G., Mewes H.W., Mannhaupt G.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA German Neurospora genome project;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL451021; CAC18297.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 815 AA; 88080 MW; 79A39BE8C6232D7D CRC64;

Query Match 75.0%; Score 30; DB 3; Length 815;
 Best Local Similarity 75.0%; Pred. No. 3.7e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
 Db 424 RLSSMVKK 431
 ||:|||||

RESULT 18
 Q9UXB1
 ID Q9UXB1 PRELIMINARY; PRT; 986 AA.
 AC Q9UXB1;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE ISOLEUCYL-TRNA SYNTHETASE (ISOLEUCINE-TRNA SYNTHETASE) (ILES)
 DE (EC 6.1.1.5).
 GN ORF-C10.008 OR ILES.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RX MEDLINE=21332296; PubMed=11427726;
 RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
 RA Aweys M.J., Chan-Weiher C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erasus G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
 RT "The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 DR EMBL; Y18930; CAB57581.1; -.
 DR EMBL; AE006697; AAK41020.1; -.
 DR HSP; P56690; IILE.
 DR InterPro; IPR002300; tRNA-synt_1a.

DR InterPro; IPR002301; tRNA-synt_ile.
 DR InterPro; IPR002303; tRNA-synt_val.
 DR Pfam; PF00133; tRNA-synt_1; 1.
 DR PRINTS; PR00984; TRNASYNTHILE.
 DR Aminoacyl-tRNA synthetase; Ligase; Complete proteome.
 SQ SEQUENCE 986 AA; 114713 MW; 0177A3C31280F5A1 CRC64;

Query Match 75.0%; Score 30; DB 17; Length 986;
 Best Local Similarity 66.7%; Pred. No. 4.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 | :||:||||
 Db 627 RFNSMLKKV 635

RESULT 19
 Q9PAGO PRELIMINARY; PRT; 1167 AA.
 AC Q9PAGO;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CHROMOSOME SEGREGATION PROTEIN.
 GN XF2558.

OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;

RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Queglio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 406:151-159(2000).
 DR EMBL; AE004063; AAF85355.1; -;
 DR InterPro; IPR003439; ABC.transportr.
 DR InterPro; IPR003405; SMC_C.
 DR InterPro; IPR003395; SMC_N.
 DR Pfam; PF02483; SMC_C; 1.
 DR Pfam; PF02463; SMC_N; 1.
 KW Complete proteome.
 SQ SEQUENCE 1167 AA; 131049 MW; BDF50AFDD833C190 CRC64;

Query Match 75.0%; Score 30; DB 16; Length 1167;
 Best Local Similarity 66.7%; Pred. No. 5.3e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 | :||:||||
 Db 1109 RLASVKKEM 1117

RESULT 20
 Q9N8K7 PRELIMINARY; PRT; 1768 AA.
 AC Q9N8K7;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 197.6 KDA PROTEIN.
 GN CHR1_292.

OS Trypanosoma brucei.
 OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
 OX NCBI_TaxID=5691;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=TREU927;
 RA Hall N., Bowman S., Quail M., Ivens A.C., Kay M.P., Bray-Allen S.,
 RA Lennard N.J., Clark L.N., Harris B.R., Melville S., Lawson D.,
 RA Gerrard C., Rajandream M.A., Barrall B.G.;
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL359782; CAB95536.1; -;
 DR InterPro; IPR001664; IF; -;
 DR InterPro; IPR003862; sub_transporter.
 DR PROSITE; PS00226; IF; UNKNOWN_1.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 1768 AA; 197642 MW; 4F91AA4DE9DE200F CRC64;

Query Match 75.0%; Score 30; DB 5; Length 1768;
 Best Local Similarity 66.7%; Pred. No. 8e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 | :||:||||
 Db 1424 RLNSVLKKV 1432

RESULT 21
 Q9GZG6 PRELIMINARY; PRT; 1887 AA.
 ID Q9GZG6
 AC Q9GZG6;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL 214.2 KDA PROTEIN.
 GN F56A6.2.

OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RX MEDLINE=99069613; PubMed=9851916;
 RA None;

RT "Genome sequence of the nematode C. elegans: a platform for
 RT investigating biology. The C. elegans Sequencing Consortium.";
 RL Science 282:2012-2018(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RC Murray J., Rohlfling T., O'Neal D., Wilson R.;
 RT "The sequence of C. elegans cosmid F56A6.";
 RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BRISTOL N2;
 RC Waterston R.;

RT "Direct Submission";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF067217; AAF99977.1; -;
DR HSP: P08799; 1MND;
DR InterPro: IPR002219; DAG_PE-bind.
DR InterPro: IPR000048; IQ;
DR InterPro: IPR001609; myosin_head.
DR InterPro: IPR001159; RA;
DR Pfam: PF00063; myosin_head; 2.
DR Pfam: PF00620; RhoGAP; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODom: PD000355; myosin_head; 2.
DR SMART: SM00109; C1; 2.
DR SMART: SM00115; IQ; 4.
DR SMART: SM00242; MYSC; 1.
DR SMART: SM00334; RA; 1.
DR SMART: SM00324; RhoGAP; 1.
DR PROSITE: PS0081; DAG_PE_BIND_DOM_2; 3.
KW Hypothetical protein.
SQ SEQUENCE 1887 AA; 214207 MW; 992DDE7F409D2D84 CRC64;

Query Match 75.0%; Score 30; DB 5; Length 1887;
Best Local Similarity 85.7%; Pred. No. 8.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 SSMVKV 9
||:||||
Db 1460 SSMIRKV 1466

RESULT 22
Q40388 PRELIMINARY; PRT; 118 AA.
AC Q40388;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PHYTOENE DEHYDROGENASE (FRAGMENT).
OS Nicotiana benthamiana.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asterales; euasterids I; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4100;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LEAF;
RX MEDLINE=95183529; PubMed=7878039;
RA Kumagai M.H., Donson J., della-Cioppa G., Harvey D., Hanley K.,
RA Grill L.K.;
RT "Cytoplasmic inhibition of carotenoid biosynthesis with virus-derived
RT RNA";
RL Proc. Natl. Acad. Sci. U.S.A. 92:1679-1683(1995).
DR EMBL: U19262; AAC48983.1; -;
FT NON_TER 118
SQ SEQUENCE 118 AA; 13356 MW; BCA393FD28B26DB1 CRC64;

Query Match 72.5%; Score 29; DB 10; Length 118;
Best Local Similarity 55.6%; Pred. No. 91;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLSSMKV 9
||:||||
Db 47 RLNSRIKKI 55

RESULT 23
Q9FG20 PRELIMINARY; PRT; 242 AA.
AC Q9FG20;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE SIMILARITY TO MADS-BOX TRANSCRIPTION FACTOR.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OX eurosids II; Brassicales; Brassicaceae; Arabidopsids.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP002543; BAB11399.1; -;
DR InterPro: IPR002100; MADS-box.
DR Pfam: PF00319; SRF-TF; 1.
DR PRINTS: PR00404; MADSDOMAIN.
DR SMART: SM00432; MADS; 1.
DR PROSITE: PS0066; MADS_BOX_2; 1.
SQ SEQUENCE 242 AA; 27298 MW; 51A14965211937B1 CRC64;

Query Match 72.5%; Score 29; DB 10; Length 242;
Best Local Similarity 55.6%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLSSMKV 9
||:||||
Db 24 RLGLIRKV 32

RESULT 24
Q97F67 PRELIMINARY; PRT; 285 AA.
AC Q97F67;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE S-ADENOSYLMETHIONINE-DEPENDENT METHYLTRANSFERASE, HEMK
DE ORTHOLOG.
GN CAC2885.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007786; AAK80828.1; -;
DR InterPro: IPR002052; N6_Mtase.
DR InterPro: IPR000051; SAM_bind.
DR PROSITE: PS00092; N6_MTASE; UNKNOWN_1.
KW Transferase; Methyltransferase; Complete proteome.
SQ SEQUENCE 285 AA; 32658 MW; 41F6E6D2859F538D CRC64;

Query Match 72.5%; Score 29; DB 16; Length 285;
Best Local Similarity 66.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLSSMKV 9
||:||||
Db 226 RITSMKKV 234

RESULT 25
O26946


```
ID O26946 PRELIMINARY; PRT; 336 AA.
AC O26946;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE GTP-BINDING PROTEIN, GTP1/OBG FAMILY.
GN MTH858.
OS Methanothermobacter thermautotrophicus.
OC Archaea: Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DELTA H;
RX MEDLINE=98037514; PubMed=9371463;
RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum
RT deltaH: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL; AE000862; AAB85356.1; -.
DR InterPro: IPR000765; GTP1_OBG.
DR Pfam: PF01018; GTP1_OBG; 1.
KW Complete proteome.
SQ SEQUENCE 336 AA; 38034 MW; B7A538D144CF87D CRC64;

Query Match 72.5%; Score 29; DB 17; Length 336;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMKVK 9
Db 130 RISSVIKRI 138
I:||||:

RESULT 26
Q97ZT5
ID Q97ZT5 PRELIMINARY; PRT; 345 AA.
AC Q97ZT5;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE THREONINE SYNTHASE (THRC-1) (EC 4.2.99.2).
GN THRC-1.
OS Sulfolobus solfataricus.
OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Awayez M.J., Chan-Welher C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sengen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR EMBL; AE006680; AAK40813.1; -.
DR InterPro: IPR001926; PALP.
DR Pfam: PF00291; PALP; 1.
KW Lyase; Complete proteome.
SQ SEQUENCE 345 AA; 38155 MW; 7D23D6D3A25B1FF8 CRC64;

Query Match 72.5%; Score 29; DB 17; Length 345;
Best Local Similarity 75.0%; Pred. No. 2.6e+02;
```

```
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMKVK 9
Db 2 ISSMKVK 9
I:||||:

RESULT 27
Q97628
ID Q97628 PRELIMINARY; PRT; 347 AA.
AC Q97628;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN ST0041.
GN ST0041.
OS Sulfolobus tokodaii.
OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AF000981; BAB64996.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 347 AA; 40165 MW; 00232EDF1777F9ED CRC64;

Query Match 72.5%; Score 29; DB 17; Length 347;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMKVK 9
Db 170 LNSMKVK 177
I:||||:

RESULT 28
Q972T3
ID Q972T3 PRELIMINARY; PRT; 347 AA.
AC Q972T3;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN ST1050.
GN ST1050.
OS Sulfolobus tokodaii.
OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AF000984; BAB66080.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 347 AA; 40217 MW; 06698A9E3F359407 CRC64;
```

Query Match 72.5%; Score 29; DB 17; Length 347;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKKV 9

I:|||||

Db 170 LNSMVKV 177

RESULT 29

ID Q972N5 PRELIMINARY; PRT; 347 AA.
AC Q972N5;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN ST1098.
GN ST1098.
OS Sulfolobus tokodaii.
OC Archaea: Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000985; BAB66130.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 347 AA; 40153 MW; 00232EC49D5F40ED CRC64;

Query Match 72.5%; Score 29; DB 17; Length 347;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKKV 9

I:|||||

Db 170 LNSMVKV 177

RESULT 30

ID Q9HR20 PRELIMINARY; PRT; 349 AA.
AC Q9HR20;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE VNG0903C.
GN VNG0903C.
OS Halobacterium sp. (strain NRC-1).
OC Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;
OX Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebbardt H., Lowe T.W., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005028; AAG19338.1; -;
DR InterPro: IPR003760; Bmp.
DR Pfam: PF02608; Bmp; 1.
KW Complete proteome.
SQ SEQUENCE 349 AA; 36438 MW; BE7415AFF330C206 CRC64;

Query Match 72.5%; Score 29; DB 17; Length 349;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKKV 9

I:|||||

Db 273 LASMVKV 280

RESULT 31

ID Q97L60 PRELIMINARY; PRT; 357 AA.
AC Q97L60;
DT 01-OCT-2001 (TRENBLrel. 18, Created)
DT 01-OCT-2001 (TRENBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE PREDICTED LIPOPROTEIN, MED/BMP FAMILY.
GN CAC0702.
OS Clostridium acetobutylicum.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OX Clostridium
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A7CC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007585; AAK78679.1; -;
DR InterPro: IPR003760; Bmp.
DR Pfam: PF02608; Bmp; 1.
KW Lipoprotein; Complete proteome.
SQ SEQUENCE 357 AA; 38354 MW; 2A3EF3AED879B686 CRC64;

Query Match 72.5%; Score 29; DB 16; Length 357;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKKV 9

I:|||||

Db 270 LASMVKV 277

RESULT 32

ID Q9TJR6 PRELIMINARY; PRT; 359 AA.
AC Q9TJR6;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE CELL DIVISION INHIBITOR MIND.
GN MIND.
OS Prototheca wickerhamii.
OC Chloroplast.
OC Eukaryota: Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Prototheca.
OX NCBI_TaxID=3111;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=263-11;
RA Knauf U., Hachtel W.;

RT "A 22 kb fragment of the 53 kb plastid genome of the colourless alga
 RT Prototheca wickerhamii containing atp-, rpl-, rps-, rrn-, and trn-
 RT genes".
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AJ245643; CAB53105.1; -
 DR InterPro; IPR000707; Para.
 DR Pfam; PF00991; Para; 1.
 KW Chloroplast.
 SQ SEQUENCE 359 AA; 40804 MW; E550EAF50BC0A51E CRC64;

 Query Match 72.5%; Score 29; DB 8; Length 359;
 Best Local Similarity 75.0%; Pred. No. 2.7e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 1 RLSSMVKK 8
 |||:|:|
 Db 32 RLSSNVKK 39

 RESULT 33
 Q9RSQ4 PRELIMINARY; PRT; 379 AA.
 AC Q9RSQ4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE MEMBRANE LIPOPROTEIN.
 GN DR2070.
 OS Deinococcus radiodurans.
 OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
 OX NCBI_TaxID=1299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RI;
 RX MEDLINE=20036896; PubMed=10567266;
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
 RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
 RA Fraser C.M.;
 RT "Genome sequence of the radioresistant bacterium Deinococcus
 RT radiodurans RI".
 RL Science 286:1571-1577(1999).
 DR EMBL; AE002043; AAF11620.1; -
 DR TIGR; DR2070; -
 DR InterPro; IPR003760; Bmp.
 DR Pfam; PF02608; Bmp; 1.
 KW Lipoprotein; Complete proteome.
 SQ SEQUENCE 379 AA; 40267 MW; 846CEFAFB2CE2BF3 CRC64;

 Query Match 72.5%; Score 29; DB 16; Length 379;
 Best Local Similarity 75.0%; Pred. No. 2.9e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

 QY 2 LSSMVKKV 9
 |:|:|:|:|
 Db 305 LTSNVKKV 312

 RESULT 34
 Q974I0 PRELIMINARY; PRT; 434 AA.
 AC Q974I0;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN ST0678.
 GN ST0678.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=111955;

RN SEQUENCE FROM N.A.
 RP STRAIN=JCM 10545 / 7;
 RX PubMed=11572479;
 RA Kawarayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermoacidophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain7".
 RL DNA Res. 8:123-140(2001).
 DR EMBL; AP000983; BAB65678.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 434 AA; 50128 MW; F3165940F22C32C2 CRC64;

 Query Match 72.5%; Score 29; DB 17; Length 434;
 Best Local Similarity 75.0%; Pred. No. 3.3e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

 QY 2 LSSMVKKV 9
 |:|:|:|:|
 Db 257 LNSWVRKV 264

 RESULT 35
 Q972H0 PRELIMINARY; PRT; 434 AA.
 AC Q972H0;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN ST1160.
 GN ST1160.
 OS Sulfolobus tokodaii.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=111955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JCM 10545 / 7;
 RX PubMed=11572479;
 RA Kawarayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
 RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
 RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
 RA Oshima T., Kikuchi H.;
 RT "Complete genome sequence of an aerobic thermoacidophilic
 RT Crenarchaeon, Sulfolobus tokodaii strain7".
 RL DNA Res. 8:123-140(2001).
 DR EMBL; AP000985; BAB66196.1; -
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 434 AA; 49954 MW; CC0D84D4855F98BE CRC64;

 Query Match 72.5%; Score 29; DB 17; Length 434;
 Best Local Similarity 75.0%; Pred. No. 3.3e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

 QY 2 LSSMVKKV 9
 |:|:|:|:|
 Db 257 LNSWVRKV 264

 RESULT 36
 Q96ZC6 PRELIMINARY; PRT; 434 AA.
 AC Q96ZC6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN ST1906.
 GN ST1906.

```

OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AF000988; BAB66998.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 434 AA; 50059 MW; D91F7BED9ADD0413 CRC64;

Query Match 72.5%; Score 29; DB 17; Length 434;
Best Local Similarity 75.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSMVKKV 9
Db 257 LNSMVRKV 264
I:|||||

RESULT 37
Q96XG3 PRELIMINARY; PRT; 436 AA.
AC Q96XG3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN ST2552.
GN ST2552.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL; AP000990; BAB67664.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 436 AA; 51017 MW; D5BE4D13AE8CB6D5 CRC64;

Query Match 72.5%; Score 29; DB 17; Length 436;
Best Local Similarity 75.0%; Pred. No. 3.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSMVKKV 9
Db 259 LNSMVRKV 266
I:|||||

RESULT 38
Q9HJ78 PRELIMINARY; PRT; 469 AA.
AC Q9HJ78;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

```

```

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MERCURIC REDUCTASE RELATED PROTEIN.
GN TAJ341.
OS Thermoplasma acidophilum.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=2303;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DSM 1728;
RX MEDLINE=20479972; PubMed=11029001;
RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
RA Meves H.-W., Frishman D., Stocker S., Lupas.A.N., Baumeister W.;
RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
RT acidophilum.";
RL Nature 407:508-513(2000).
CC -I- COFACTOR: FAD (BY SIMILARITY).
DR EMBL; AL445067; CAC12462.1; -.
DR HSSP; P11959; 1EBD.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000815; Hq_reductase.
DR InterPro; IPR000205; NAD_binding.
DR InterPro; IPR001230; Prenyltn.
DR InterPro; IPR001003; Pyridine_redox_2.
DR InterPro; IPR001100; pyr_redox.
DR InterPro; IPR004099; pyr_redox_dim.
DR Pfam; PF00070; pyr_redox; 1.
DR Pfam; PF02852; pyr_redox_dim; 1.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00945; HGRDTASE.
DR PRINTS; PR00411; PNDRDTASEII.
DR PROSITE; PS00294; PRENYLATION; UNKNOWN_1.
DR PROSITE; PS00076; PYRIDINE_REDOX_1; 1.
DR Complete proteome; FAD; Flavoprotein; Oxidoreductase.
SQ SEQUENCE 469 AA; 50583 MW; D72FE339815B5B51 CRC64;

Query Match 72.5%; Score 29; DB 17; Length 469;
Best Local Similarity 87.5%; Pred. No. 3.6e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LSSMVKKV 9
Db 233 LSSMVKDV 240
I:|||||

RESULT 39
Q9X1Y5 PRELIMINARY; PRT; 553 AA.
AC Q9X1Y5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE HYPOTHETICAL 61.3 KDA PROTEIN.
GN TMI652.
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
RA Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
RA McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
RA Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
RA Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
RA Salzberg S.L., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from
RT genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).
DR EMBL; AE001807; AAD36719.1; -.
DR TIGR; TMI652; -.

```

KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 553 AA: 61270 MW; DA5F5C9625BAC6A0 CRC64;

Query Match 72.5%; Score 29; DB 16; Length 553;
Best Local Similarity 66.7%; Pred. No. 4.2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
II IIII

Db 269 RLDEWVKRV 277

RESULT 40

Q9PPR7 PRELIMINARY; PRT; 578 AA.
AC Q9PPR7:
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN UU572.
GN UU572.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes.
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Caselli G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
urealyticum".
RL Nature 407:757-762(2000).
DR EMBL; AE002156; AAF30986.1; --
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 578 AA: 67401 MW; 9E9758B1F0AA8724 CRC64;

Query Match 72.5%; Score 29; DB 16; Length 578;
Best Local Similarity 55.6%; Pred. No. 4.4e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
II IIII

Db 98 RLNTVVKKI 106

RESULT 41

Q965D0 PRELIMINARY; PRT; 593 AA.
AC Q965D0:
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE ABC TRANSPORTER.
GN ABC1.
OS Theileria parva.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC Theileria.
OX NCBI_TaxID=5875;
RN [1]
RP SEQUENCE FROM N.A.
RA Kibe M.K., Macklin M., Gobright E., Bishop R., Urakawa T.,
RA ole-Moirot O.K.;
RT "Characterization and expression of single domain ATP-binding cassette
protein homologs of the lymphocyte-transforming protozoan parasite
Theileria parva".
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF255047; AAK72899.1; --
SQ SEQUENCE 593 AA: 66465 MW; 03465AAB76557859 CRC64;

Query Match 72.5%; Score 29; DB 5; Length 593;
Best Local Similarity 75.0%; Pred. No. 4.5e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 RLSSMVKK 8
:IIIIII
Db 173 QLSSMVKKR 180

RESULT 42

Q9XU08 PRELIMINARY; PRT; 656 AA.
AC Q9XU08:
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE K05C4.5 PROTEIN.
GN K05C4.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Harris B.R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology".
RL Science 282:2012-2018(1998).
CC -!- SIMILARITY: CONTAINS 4 WD REPEATS (TRP-ASP DOMAINS).
DR EMBL; Z81564; CAB04569.1; --
DR InterPro; IPR001860; WD40.
DR Pfam; PF00400; WD40; 4.
KW Repeat; WD repeat.
SQ SEQUENCE 656 AA: 72791 MW; 0E79CE50322F48D5 CRC64;

Query Match 72.5%; Score 29; DB 5; Length 656;
Best Local Similarity 55.6%; Pred. No. 5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
:IIIIII

Db 366 KISPMVKKI 374

RESULT 43

O07430 PRELIMINARY; PRT; 691 AA.
AC O07430:
DT 01-JUL-1997 (TRENBLrel. 04, Created)
DT 01-JUL-1997 (TRENBLrel. 04, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE BETA-GLUCOSIDASE.
GN BGLS OR RV0186 OR MTCI28.25B.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Horsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skellon S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the

```

RT complete genome sequence." ;
RL Nature 393:537-544 (1998).
DR EMBL: Z97050; CAB09737.1; -.
DR TubercuList: Rv0186; -.
DR InterPro: IPR001764; Glyco_hydro_3.
DR InterPro: IPR002772; Glyco_hydro_3C.
DR Pfam: PF00933; Glyco_hydro_3; 1.
DR Pfam: PF01915; Glyco_hydro_3_C; 1.
DR PRINTS: PR00133; GLHYDRLASE3.
KW Complete proteome.
SQ SEQUENCE 691 AA; 73520 MW; B700C8C37F24E1F7 CRC64;

Query Match 72.5%; Score 29; DB 16; Length 691;
Best Local Similarity 55.6%; Pred. NO. 5.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLSSMVKKV 9
DB 282 RLSDMVRRI 290

RESULT 44
Q32S82 PRELIMINARY; PRT; 849 AA.
AC Q95Z82
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE PUTATIVE ATP-DEPENDENT HELICASE PROTEIN (EC 3.6.1.1).
GN SMC02237.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21366234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Burnmeser J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Ransperger U., Surzycki R., Thebault P., Vandenbol M., Purnelle B.,
RA Vorhoefer F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti." ;
RL Science 293:668-672(2001).
DR EMBL: AL591784; CAC45113.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 849 AA; 93953 MW; 298D8094FBC39A69 CRC64;

Query Match 72.5%; Score 29; DB 16; Length 849;
Best Local Similarity 55.6%; Pred. NO. 6.4e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLSSMVKKV 9
DB 785 RLGSMLKRI 793

RESULT 45
Q95ZA8 PRELIMINARY; PRT; 1165 AA.
ID Q95ZA8
AC Q95ZA8;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE POSSIBLE RNA BINDING PROTEIN.
GN L3747.04.
OS Leishmania major.

```

```

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Rieger M., Fuchs M., Gabel C., Mueller-Auer S., Schaefer M.,
RA Ivens A.C., Quail M., Rajandream M.A., Barrell B.G.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome." ;
RL Genome Res. 8:135-145(1998).
DR EMBL: AL596273; CAC44733.1; -.
SQ SEQUENCE 1165 AA; 122865 MW; 1281663DB835695B CRC64;

Query Match 72.5%; Score 29; DB 5; Length 1165;
Best Local Similarity 66.7%; Pred. NO. 8.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLSSMVKKV 9
DB 694 RLSSMFKRL 702

Search completed: November 6, 2002, 12:11:58
Job time : 22.4444 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 11 Seconds
(without alignments)
78.619 Million cell updates/sec

Title: US-09-905-083-32
Perfect score: 40
Sequence: 1 RLSSMVKKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	253	A53968	serine proteinase
2	33	82.5	282	T35294	probable endo aliph
3	33	82.5	372	S20056	para-hydroxybenzoa
4	33	82.5	560	S50401	hypothetical prote
5	32	80.0	350	C70009	ABC transporter (l
6	32	80.0	351	D90264	biotin synthase (b
7	32	80.0	358	T38914	para-hydroxybenzoa
8	32	80.0	589	A12777	arginine--trNA lig
9	32	80.0	664	T50316	hypothetical Armad
10	31	77.5	912	H85470	hypothetical prote
11	31	77.5	925	T02473	hypothetical prote
12	30	75.0	239	T02473	hypothetical prote
13	30	75.0	240	T32363	hypothetical prote
14	30	75.0	318	T36519	taipoxin-associate
15	30	75.0	319	JC5402	vitamin D receptor
16	30	75.0	373	G75073	cytochrome P450 li
17	30	75.0	457	D85429	glycine hydroxymet
18	30	75.0	532	T01759	hemagglutinin homo
19	30	75.0	647	A49218	transcription fact
20	30	75.0	833	A47528	isoleucine--trNA sy
21	30	75.0	986	E90220	chromosome segrega
22	30	75.0	1167	A82543	hypothetical prote
23	30	75.0	1846	T33079	TATA-binding prote
24	29	72.5	253	JC5511	S-adenosylmethioni
25	29	72.5	285	A97255	GTP-binding protei
26	29	72.5	336	E69214	Deinococcus radiol
27	29	72.5	337	B97374	membrane lipoprote
28	29	72.5	337	AH2591	threonine synthase
29	29	72.5	345	F90194	

30	29	72.5	349	2	F84246	hypothetical prote
31	29	72.5	357	2	D96986	probable lipoprote
32	29	72.5	379	2	H75318	membrane lipoprote
33	29	72.5	553	2	A72228	hypothetical prote
34	29	72.5	571	2	S65060	phytoene desaturas
35	29	72.5	578	2	H82872	hypothetical prote
36	29	72.5	582	2	S29314	phytoene dehydroge
37	29	72.5	583	2	A45381	hypothetical prote
38	29	72.5	656	2	T33338	probable beta-gluc
39	29	72.5	691	2	E70906	hypothetical prote
40	29	72.5	839	2	S62963	hypothetical prote
41	29	72.5	1360	2	F96596	hypothetical prote
42	29	72.5	4077	2	T17484	hypothetical prote
43	29	72.5	4589	2	T14914	dynein beta heavy
44	28	70.0	41	2	G71376	hypothetical prote
45	28	70.0	63	2	AH2565	hypothetical prote

ALIGNMENTS

RESULT 1

A53968
serine proteinase SCCE precursor - human
N:Alternate names: stratum corneum chymotryptic enzyme
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
C:Accession: A53968
R:Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlsteinn, A.; Egelrud, J. Biol. Chem. 269, 19420-19426, 1994
A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic en
A:Reference number: A53968; MUID: 94308225
A:Accession: A53968
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253 <HAN>
A:Cross-references: GB:L33404; NID:g521214; PIDN:AAC37551.1; PID:g532504
C:Genetics:
A:Gene: GDB:PRSS6; SCCE
A:Cross-references: GDB:377730
A:Map position: 7q35-7q35
C:Superfamily: trypsin; trypsin homology
F:30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
Db 123 RLSSMVKKV 131
|||||

RESULT 2

T35294
probable endo alpha-1,4 polygalactosaminidase - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.P.
C:Accession: T35294
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21574
A:Accession: T35294
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-282 <SEE>
A:Cross-references: EMBL:AL056872; PIDN:CAB51262.1; GSPDB:GN00070; SCOEDB:SC5F7.23c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOEDB:SC5F7.23c

Query Match 82.5%; Score 33; DB 2; Length 282;
Best Local Similarity 87.5%; Pred. No. 13;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8

|||||

Db 265 RLSSMLKK 272

RESULT 3

S20056 para-hydroxybenzoate polyprenyltransferase (EC 2.5.1.-) precursor - yeast (Saccharomyces

N:Alternate names: 4-hydroxybenzoate hexaprenyltransferase; protein N3419; protein YNR04

C:Species: Saccharomyces cerevisiae

C:Date: 23-Apr-1993 #sequence_revision 23-Apr-1993 #text_change 21-Jul-2000

C:Accession: S20056; S63372; JC2317

R:Ashby, M.N.; Kutsunai, S.Y.; Ackerman, S.; Tzagoloff, A.; Edwards, P.A.

J. Biol. Chem. 267, 4128-4136, 1992

A:Title: COQ2 is a candidate for the structural gene encoding para-hydroxybenzoate:polyp

A:Reference number: S20056

A:Accession: S20056

A:Molecule type: DNA

A:Residues: 1-372 <ASH>

A:Cross-references: EMBL:M81698; NID:gl71252; PIDN:AAA34507.1; PID:gl71254

R:Pohl, T.M. submitted to the Protein Sequence Database, April 1996

A:Reference number: S63372

A:Accession: S63372

A:Molecule type: DNA

A:Residues: 1-372 <POH>

A:Cross-references: EMBL:Z71656; NID:gl302546; PID:e239832; PID:gl302547; MIPS:YNR041c

A:Experimental source: strain S288C

C:Genetics:

A:Gene: SGD:COQ2

A:Cross-references: SGD:S0005324; MIPS:YNR041c

A:Map position: 14R

C:Keywords: isoprenoid biosynthesis; mitochondrion; transferase; transmembrane protein

F:1-22/Domain: transit peptide (mitochondrion) #status predicted <TRP>

F:117-133/Domain: transmembrane #status predicted <TM1>

F:172-188/Domain: transmembrane #status predicted <TM2>

F:203-225/Region: polyprenyl diphosphate binding #status predicted

Query Match

Best Local Similarity 82.5%; Score 33; DB 2; Length 372;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 9

|||||

Db 326 RLFSMLKK 334

RESULT 4

S50401

hypothetical protein YMR145c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein YW9375.14c

C:Species: Saccharomyces cerevisiae

C:Date: 13-Jan-1995 #sequence_revision 24-Feb-1995 #text_change 05-Nov-1999

C:Accession: S50401

R:Badcock, K.; Churcher, C.

submitted to the EMBL Data Library, December 1994

A:Reference number: S50398

A:Accession: S50401

A:Molecule type: DNA

A:Residues: 1-560 <BAD>

A:Cross-references: EMBL:Z47071; NID:g606429; PIDN:CAA87359.1; PID:g606443; GSPDB:GN0001

C:Genetics:

A:Gene: MIPS:YMR145c

A:Map position: 13R

C:Superfamily: NADH dehydrogenase

Query Match

Best Local Similarity 82.5%; Score 33; DB 2; Length 560;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 9

Db 347 RLKTMVKV 355

|||||

RESULT 5

C70009

ABC transporter (lipoprotein) homolog yufN - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000

C:Accession: C70009

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.: Bron, S.; Brouillet, S.; Bruchli, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.: Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033

A:Accession: C70009

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-350 <KUN>

A:Cross-references: GB:Z99120; GB:AL009126; NID:g2635613; PIDN:CABL5143.1; PID:g26356

A:Experimental source: strain 168

C:Genetics:

A:Gene: yufN

C:Superfamily: ABC transporter yufN

Query Match

Best Local Similarity 80.0%; Score 32; DB 1; Length 350;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKKV 9

|||||

Db 269 LTSVMKKV 276

RESULT 6

D90264

biotin synthase (bioB) [imported] - Sulfolobus solfataricus

C:Species: Sulfolobus solfataricus

C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001

C:Accession: D90264

R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Awayez, M.J.; Ch

arrett, R.A.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder

submitted to GenBank, April 2001

A:Description: Sulfolobus solfataricus complete genome.

A:Reference number: A99139

A:Accession: D90264

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-351 <KUR>

A:Cross-references: GB:AF006641; NID:gl3814305; PIDN:AAK41371.1; GSPDB:GN00155

C:Genetics:

A:Gene: bioB

Query Match

Best Local Similarity 80.0%; Score 32; DB 2; Length 351;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

|||||

Db 341 RLDSLIRKV 349

RESULT 7

T38914
para-hydroxybenzoate--polyprenyltransferase (EC 2.5.1.1-) precursor, mitochondrial - fliss
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T38914
R:Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL Data Library, February 1996
A:Reference number: Z21817
A:Accession: T38914
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-358 <PEA>
A:Cross-references: EMBL:Z69728; PIDN:CAA93575.1; GSPDB:GN00066; SPDB:SPAC56F8.04c
A:Experimental source: strain 972h-; cosmid c56F8
C:Genetics:
A:Gene: SPDB:SPAC56F8.04c
A:Map position: 1
A:Genome: nuclear
C:Keywords: mitochondrion; transferase

Query Match 80.0%; Score 32; DB 2; Length 358;
Best Local Similarity 77.8%; Pred. No. 27;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 311 RLSSMIYKV 319

RESULT 8

A71277
arginine--tRNA ligase (EC 6.1.1.19) (args) - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: A71277
R:Fraser, C.M.; Norris, S.J.; Weinstein, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McD
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 373-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: A71277
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-589 <COL>
A:Cross-references: GB:AE001253; GB:AE000520; NID:g3323133; PIDN:AAC65797.1; PID:g332314
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0831
C:Superfamily: Bacillus arginine--tRNA ligase
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 80.0%; Score 32; DB 2; Length 589;
Best Local Similarity 66.7%; Pred. No. 44;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 473 RLSSLLKKV 481

RESULT 9

T50316
hypothetical Armadillo/beta-catenin domain protein [imported] - flission yeast (Schizosac
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jun-2000
C:Accession: T50316
R:McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.; Cadieu, E.; Lelaure, V.; Galibert, F
submitted to the EMBL Data Library, January 2000
A:Reference number: Z25061
A:Accession: T50316

A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-664 <WCD>

A:Cross-references: EMBL:AL136536; PIDN:CAB66447.1; GSPDB:GN00067; SPDB:SPBC1703.03c
A:Experimental source: strain 972h(-); cosmid cl1703
C:Genetics:
A:Gene: SPDB:SPBC1703.03c
A:Map position: 2
A:Introns: 43/3

Query Match 80.0%; Score 32; DB 2; Length 664;
Best Local Similarity 66.7%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 624 RLKNNVKKI 632

RESULT 10

H85470
hypothetical protein AT4g39750 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001
C:Accession: H85470
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp
Nature 402, 769-777, 1999
A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: H85470
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-912 <STO>
A:Cross-references: GB:NC_001268; NID:g7270959; PIDN:CAB80638.1; GSPDB:GN00140
C:Genetics:
A:Gene: AT4g39750
A:Map position: 4

Query Match 77.5%; Score 31; DB 2; Length 912;
Best Local Similarity 55.6%; Pred. No. 11e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 403 RFSSMIRKI 411

RESULT 11

T05012
hypothetical protein T19P19.140 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C:Accession: T05012
R:Bevan, M.; Monfort, A.; Casacuberta, E.; Puigdomenech, P.; Hoheisel, J.; Mewes, H.W
submitted to the Protein Sequence Database, April 1998
A:Reference number: Z15394
A:Accession: T05012
A:Molecule type: DNA
A:Residues: 1-925 <BEV>
A:Cross-references: EMBL:AL022605
A:Experimental source: cultivar Columbia; BAC clone T19P19
C:Genetics:
A:Map position: 4
A:Introns: 142/3; 193/1; 551/1
A:Note: T19P19.140

Query Match 77.5%; Score 31; DB 2; Length 925;
Best Local Similarity 55.6%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 416 RFSSMIRKI 424

RESULT 12

T02473
hypothetical protein At2g45740 [imported] - Arabidopsis thaliana
N:Alternate names: hypothetical protein F4118.28
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
C:Accession: T02473; C84894
R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul
submitted to the EMBL Data Library, August 1998
A:Description: Arabidopsis thaliana chromosome II BAC F4118 genomic sequence.
A:Reference number: Z14674
A:Accession: T02473
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-239 <ROU>
A:Cross-references: EMBL:AC004665; NID:g3386593; PID:g3386621
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487
A:Accession: T02473
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-239 <STO>
A:Cross-references: GB:AE002093; NID:g3386621; PID:AC28551.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g45740; F4118.28
A:Map position: 2
A:Introns: 65/3; 90/3; 115/3; 142/3; 167/3

Query Match 75.0%; Score 30; DB 2; Length 239;

Best Local Similarity 66.7%; Pred. NO. 52;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 148 RLSSMKKI 156

RESULT 13

T32363
hypothetical protein C08E3.12 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 29-Oct-1999
C:Accession: T32363
R:Miller, N.; Kramer, J.; Keppler, D.
submitted to the EMBL Data Library, September 1997
A:Description: The sequence of C. elegans cosmid C08E3.
A:Reference number: Z21155
A:Accession: T32363
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-240 <WIL>
A:Cross-references: EMBL:AF025457; PID:AB70970.1; GSPDB:GN00020; CESP:C08E3.12
A:Experimental source: strain Bristol N2; clone C08E3
C:Genetics:
A:Gene: CESP:C08E3.12
A:Map position: 2
A:Introns: 167/3

Query Match 75.0%; Score 30; DB 2; Length 240;

Best Local Similarity 66.7%; Pred. NO. 53;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 193 RIPSMLKKV 201

RESULT 14

I56519
taipoxin-associated calcium binding protein-49 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 20-Sep-1999
C:Accession: I56519
R:Dodds, D.; Schlimgen, A.K.; Lu, S.Y.; Perlin, M.S.
J. Neurochem. 64, 2339-2344, 1995
A:Title: Novel reticular calcium binding protein is purified on taipoxin columns.
A:Reference number: I56519; MUID:95239201
A:Accession: I56519
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-318 <RES>
A:Cross-references: EMBL:U15734; NID:g606967; PID:AAA80197.1; PID:g606968
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: calcium binding; EF hand
F:62-94/Domain: calmodulin repeat homology <EF1>
F:120-152/Domain: calmodulin repeat homology <EF2>
F:187-219/Domain: calmodulin repeat homology <EF3>
F:228-260/Domain: calmodulin repeat homology <EF4>
F:284-296/Domain: calmodulin repeat homology <EF5>

Query Match 75.0%; Score 30; DB 2; Length 318;

Best Local Similarity 55.6%; Pred. NO. 69;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 66 RLQSIIRKI 74

RESULT 15

JC5402
vitamin D receptor associated factor 1 - mouse
N:Alternate names: ERC-55 protein
C:Species: Mus musculus (house mouse)
C:Date: 07-Jul-1997 #sequence_revision 18-Jul-1997 #text_change 18-Jun-1999
C:Accession: JC5402
R:Imai, T.; Matsuda, K.; Shimojima, T.; Hashimoto, T.; Masuhiro, Y.; Kitamoto, T.; Su
Biochem. Biophys. Res. Commun. 233, 765-769, 1997
A:Title: ERC-55, a binding protein for the papilloma virus E6 oncoprotein, specific
A:Reference number: JC5402; MUID:97312489
A:Accession: JC5402
A:Molecule type: mRNA
A:Residues: 1-319 <IMA>
C:Comment: This protein acts as a vitamin D receptor-specific cofactor modulating its
C:Superfamily: unassigned calmodulin-related proteins; calmodulin repeat homology
C:Keywords: EF hand
F:63-95/Domain: calmodulin repeat homology <EF1>
F:99-131/Domain: calmodulin repeat homology <EF2>
F:151-183/Domain: calmodulin repeat homology <EF3>
F:188-220/Domain: calmodulin repeat homology <EF4>
F:229-261/Domain: calmodulin repeat homology <EF5>
F:265-297/Domain: calmodulin repeat homology <EF6>

Query Match 75.0%; Score 30; DB 2; Length 319;

Best Local Similarity 55.6%; Pred. NO. 69;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 67 RLQSIIRKI 75

RESULT 16

G75073
hypothetical protein PAB1719 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 28-Jul-2000
C:Accession: G75073
R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: G75073
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-373 <KAW>
A:Cross-references: GB:AJ248286; GB:AL096836; NID:g5458366; PIDN:CAB49892.1; PID:g545840
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB1719
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0997

Query Match 75.0%; Score 30; DB 2; Length 373;

Best Local Similarity 77.8%; Pred. No. 80;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

II I I I I I

Db 102 RLGSYVKV 110

RESULT 17

D85429

Cytochrome P450 like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Apr-2001

C:Accession: D85429

R:anonymous: The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: D85429

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-457 <STO>

A:Cross-references: GB:NC_001268; NID:g7270586; PIDN:CAB80304.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4G36380

A:Map position: 4

C:Superfamily: Synchocystis cytochrome P450 slr0574; cytochrome P450 homology

C:Keywords: heme; iron; metalloprotein

F:396/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 75.0%; Score 30; DB 2; Length 457;

Best Local Similarity 77.8%; Pred. No. 98;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

II I I I I I

Db 199 RLKMKVKV 207

RESULT 18

T01759

glycine hydroxymethyltransferase (EC 2.1.2.1) A_IG002P16.3 - Arabidopsis thaliana

N:Alternate names: protein A_IG002P16.3

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 23-Jul-1999

C:Accession: T01759

R:Miller, N.; Beck, C.; Kramer, J.

submitted to the EMBL Data Library, June 1997

A:Description: The sequence of A. thaliana IG002P16.

A:Reference number: Z14421

A:Accession: T01759

A>Status: translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-532 <WIL>

A:Cross-references: EMBL:AF007270; NID:g2191157; PIDN:AB61045.1; PID:g2191159

C:Genetics:

A:Map position: 5

A:Introns: 29/3; 58/3; 100/3; 134/1; 175/3; 206/3; 237/3; 303/3; 338/3; 360/3; 402/3; 44

A>Note: A_IG002P16.3

C:Superfamily: glycine hydroxymethyltransferase

C:Keywords: phosphoprotein; pyridoxal phosphate; transferase

F:277/Active site: His #status predicted

F:278/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 75.0%; Score 30; DB 2; Length 532;

Best Local Similarity 87.5%; Pred. No. 1.1e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8

II I I I I I

Db 7 RLSSSVKK 14

RESULT 19

A49218

hemagglutinin homolog pmGAL.2 - Mycoplasma gallisepticum

C:Species: Mycoplasma gallisepticum

C>Date: 21-Jan-1994 #sequence_revision 18-Nov-1994 #text_change 07-Dec-1999

C:Accession: A49218

R:Markham, P.F.; Glew, M.D.; Whithear, K.G.; Walker, I.D.

Infect. Immun. 61, 903-909, 1993

A:Title: Molecular cloning of a member of the gene family that encodes pmGA, a hemagglutinin

A:Reference number: A49218; MUID:93162830

A:Accession: A49218

A>Status: preliminary

A:Molecule type: DNA; protein

A:Residues: 1-647 <MAR>

A:Cross-references: GB:S55216; NID:g265625; PIDN:AAB25397.1; PID:g265626

A:Experimental source: S6

A>Note: sequence extracted from NCBI backbone (NCBIN:125182, NCBI:P:125183)

C:Genetics:

A:Genetic code: SGC3

Query Match 75.0%; Score 30; DB 2; Length 647;

Best Local Similarity 75.0%; Pred. No. 1.4e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKKV 9

II I I I I I

Db 237 LSSIVKKI 244

RESULT 20

A47528

transcription factor MBP1 - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein D2562; protein YDL056w

C:Species: Saccharomyces cerevisiae

C>Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 20-Sep-1999

C:Accession: A47528; B47528; S59375; S67591; S62745; S37404

R:Koch, C.; Moll, T.; Neuberg, M.; Ahorn, H.; Nasmyth, K.

Science 261, 1551-1557, 1993

A:Title: A role for the transcription factors Mbpl and Swi4 in progression from G1 to

A:Reference number: A47528; MUID:93383264

A:Accession: A47528

A>Status: nucleic acid sequence not shown

A:Molecule type: DNA

A:Residues: 1-833 <KOC>

A:Cross-references: GB:X74158; NID:g402792; PIDN:CAA52271.1; PID:g402793

A:Accession: B47528

A:Molecule type: protein

A:Residues: 'X', 372-377, 'X', 379-381, 'X', 383-387 <K02>

R:Benton, B.K.; Plump, S.D.; Roos, J.; Lennarz, W.J.; Cross, F.R.

submitted to the EMBL Data Library, January 1995

A:Description: Overexpression of S. cerevisiae G1 cyclins restores viability of aigl

A:Reference number: S59374

A:Accession: S59375

A:Molecule type: DNA

A:Residues: 460-833 <BEN>

A:Cross-references: EMBL:U19608; NID:g1292897; PIDN:AAC49290.1; PID:g639493

R:Bloeker, H.; Brandt, P.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67587

A:Accession: S67591
A:Molecule type: DNA
A:Residues: 1-833 <BL0>
A:Cross-references: EMBL:Z74104; NID:gl1431054; PIDN:CAA98618.1; PID:e253002; PID:gl143105
A:Experimental source: strain S288C
R:Benton, B.K.; Plump, S.D.; Roos, J.; Lennarz, W.J.; Cross, F.R.
Curr. Genet. 29, 106-113, 1996
A:Title: Over-expression of *S. cerevisiae* G(1) cyclins restores the viability of *alg1* N-
A:Reference number: S62744; MUID:96418871
A:Accession: S62745
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 460-833 <BEW>
A:Cross-references: EMBL:U19608; NID:gl292897; PIDN:AAQ49290.1; PID:g639493
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, January 1995
C:Genetics:
A:Gene: SGD:MBPI
A:Cross-references: SGD:S0002214; MIPS:YDL056w
A:Map position: 4L
C:Superfamily: unassigned ankyrin repeat proteins; ankyrin repeat homology; EGF homology
C:Keywords: DNA binding; nucleus; transcription regulation
F:394-426/Domain: ankyrin repeat homology <ANL>

Query Match 75.0%; Score 30; DB 2; Length 833;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMWKK 8
|:|:|:|:|
Db 767 KLSSLVKK 774

RESULT 21

E90220
isoleucine-tRNA synthetase (Iles) [imported] - *Sulfolobus solfataricus*
C:Species: *Sulfolobus solfataricus*
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: E90220
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, V.; Allard, G.; Awayez, M.J.; Chan-
Jong, I.; Jeffries, A.C.; Koza, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: *Sulfolobus solfataricus* complete genome.
A:Reference number: A99139
A:Accession: E90220
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-986 <KUR>
A:Cross-references: GB:AE006641; NID:gl3813890; PIDN:AAK41020.1; GSPDB:GN00155
C:Genetics:
A:Gene: Iles
C:Superfamily: isoleucine-tRNA ligase

Query Match 75.0%; Score 30; DB 2; Length 986;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMWKKV 9
|:|:|:|:|
Db 627 RFNSMLKKV 635

RESULT 22

A82543
chromosome segregation protein XF2558 [imported] - *Xylella fastidiosa* (strain 9a5c)
C:Species: *Xylella fastidiosa*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: A82543
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen-
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below

A:Accession: A82543
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1167 <SIM>
A:Cross-references: GB:AE004063; GB:AE003849; NID:99107766; PIDN:AAF85355.1; GSPDB:GN
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorry, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuchioka, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XP2558

Query Match 75.0%; Score 30; DB 2; Length 1167;
Best Local Similarity 66.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMWKKV 9
|:|:|:|:|
Db 1109 RLASMKEM 1117

RESULT 23

T33079
hypothetical protein F56A6.2 - *Caenorhabditis elegans*
C:Species: *Caenorhabditis elegans*
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 23-Mar-2001
C:Accession: T33079
R:Murray, J.; Rohlfing, T.; O'Neal, D.; Wilson, R.
submitted to the EMBL Data Library, May 1998
A:Description: The sequence of *C. elegans* cosmid F56A6.
A:Reference number: Z21279
A:Accession: T33079
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1846 <MUR>
A:Cross-references: EMBL:AF067217; PIDN:AAC17015.1; GSPDB:GN00019; CBSP:F56A6.2
A:Experimental source: strain Bristol N2; clone F56A6
C:Genetics:
A:Gene: CBSP:F56A6.2
A:Map position: 1
A:Introns: 71/1; 165/3; 247/2; 335/2; 356/2; 389/3; 441/3; 497/3; 539/1; 607/3; 636/3
06/1; 1714/3; 1756/2
C:Superfamily: rat myosin heavy chain; myosin motor domain homology; protein kinase C

Query Match 75.0%; Score 30; DB 2; Length 1846;
Best Local Similarity 85.7%; Pred. No. 3.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 SSMVKV 9
|:|:|:|:|
Db 1419 SSMIKV 1425

RESULT 24

JC5511
TATA-binding protein-associated factor II 31 - rat
C:Species: *Rattus norvegicus* (Norway rat)
C:Date: 02-Sep-1997 #sequence_revision 05-Sep-1997 #text_change 05-Nov-1999
C:Accession: JC5511
R:Aoki, T.; Koike, T.; Nakano, T.; Shibahara, K.; Nishimura, H.; Kikuchi, H.; Honjo,
Biochem. Biophys. Res. Commun. 234, 230-234, 1997
A:Title: Rat TAFII31 gene is induced upon programmed cell death in differentiated PC1

A:Reference number: JC5511; MUID:97312553
A:Accession: JC5511
A:Molecule type: mRNA
A:Residues: 1-253 <AOK>
A:Cross-references: GB:U40188; NID:gl1103899; PIDN:AAC53201.1; PID:gl1103900
A:Experimental source: PC12 cell
C:Comment: This protein is a component of TATA-binding protein-associated factor IID, and is involved in the regulation of transcription.
C:Genetics:
A:Gene: TAFII31

Query Match 72.5%; Score 29; DB 2; Length 253;
Best Local Similarity 75.0%; Pred. No. 93;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
|||:||||
DB 123 RLKSLVK 130

RESULT 25
A97255
S-adenosylmethionine-dependent methyltransferase, HEMK ortholog [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: A97255
R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Dally, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: A97255
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-285 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80828.1; PID:gl5025933; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2885

Query Match 72.5%; Score 29; DB 2; Length 285;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 9
||:|||||
DB 226 RLSSMVKK 234

RESULT 26
E69214
GTP-binding protein, GTP1/OBG family - Methanobacterium thermoautotrophicum (strain Delta H)
C:Species: Methanobacterium thermoautotrophicum
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 02-Feb-2001
C:Accession: E69214
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwan, N.; Ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional and comparative genomics
A:Reference number: A69000; MUID:98037514
A:Accession: E69214
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <MTH>
A:Cross-references: GB:AE000862; GB:AE000666; NID:g2621943; PIDN:AAB85356.1; PID:g262195
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH858
C:Superfamily: GTP-binding protein, GTP1/OBG; translation elongation factor Tu homology
C:Keywords: GTP binding; nucleotide binding; P-loop
F:163-285/Domain: translation elongation factor Tu homology <TU>
F:169-176/Region: nucleotide-binding motif A (P-loop)
F:282-285/Region: GTP-binding NKXD motif

Query Match 72.5%; Score 29; DB 2; Length 336;
Best Local Similarity 44.4%; Pred. No. 1.2e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKK 9
||:||||:
DB 130 RLSSMVKK 138

RESULT 27
B97374
Deinococcus radiodurans membrane lipoprotein homolog [imported] - Agrobacterium tumefaciens
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: B97374
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Gold, A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markeiz, Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens
A:Reference number: A97359; PMID:11743194
A:Accession: B97374
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK85947.1; PID:gl5155002; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_200
A:Map position: circular chromosome
C:Superfamily: ABC transporter yufn

Query Match 72.5%; Score 29; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKK 9
||:||||:
DB 255 LSSMVKK 262

RESULT 28
AH2591
membrane lipoprotein [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 01-Feb-2002
C:Accession: AH2591
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; MCCI : Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AH2591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41150.1; PID:gl7738446; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu0126
A:Map position: circular chromosome
C:Superfamily: ABC transporter yufn

Query Match 72.5%; Score 29; DB 2; Length 337;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKK 9
||:||||:
DB 255 LSSMVKK 262

```
RESULT 29
F90194
threonine synthase (thrc-1) [imported] - Sulfolobus solfataricus
C:Species: Sulfolobus solfataricus
C:Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 15-Jun-2001
C:Accession: F90194
R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan-
arrett, R.A.; Ragan, M.A.; Sensen, C.W.; Van der Oost, J.
submitted to GenBank, April 2001
A:Description: Sulfolobus solfataricus complete genome.
A:Reference number: A99139
A:Accession: F90194
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-345 <KUR>
A:Cross-references: GB:AE006641; NID:gl3813649; PIDN:AAK40813.1; GSPDB:GN00155
C:Genetics:
A:Gene: thrC-1
C:Superfamily: threonine dehydratase

Query Match 72.5%; Score 29; DB 2; Length 345;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSMWKVV 9
      |||||:|
Db 2 ISSMWKEV 9

RESULT 30
F84246
hypothetical protein Vng0903c [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: F84246
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.;
Leithausen, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabil-
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: F84246
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <STO>
A:Cross-references: GB:AE004437; NID:gl0580462; PIDN:AAG19338.1; GSPDB:GN00138
C:Genetics:
A:Gene: VNG0903C
C:Superfamily: ABC transporter yufN

Query Match 72.5%; Score 29; DB 2; Length 349;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSMWKVV 9
      |||||:|
Db 273 LASMWKRV 280

RESULT 31
D96986
probable lipoprotein, Med/BMP family [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: D96986
R:Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Cl
A:Reference number: A96900; MUID:21359325; PMID:21359325
```

```
A:Accession: D96986
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-357 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK78679.1; PID:gl5023581; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC4924
C:Genetics:
A:Gene: CAC0702

Query Match 72.5%; Score 29; DB 2; Length 357;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSMWKVV 9
      |||||:|
Db 270 LTSMWKRV 277

RESULT 32
H75318
membrane lipoprotein - Deinococcus radiodurans (strain RL)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: H75318
R:White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans RL.
A:Reference number: A75250; MUID:20036896
A:Accession: H75318
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-379 <WHI>
A:Cross-references: GB:AE002043; GB:AE000513; NID:g6459859; PIDN:AAF11620.1; PID:g645
A:Experimental source: strain RL
C:Genetics:
A:Gene: DR2070
A:Map position: 1
C:Superfamily: ABC transporter yufN

Query Match 72.5%; Score 29; DB 2; Length 379;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LSSMWKVV 9
      |||||:|
Db 305 LTSMWKRV 312

RESULT 33
A72228
hypothetical protein - Thermotoga maritima (strain MSB8)
C:Species: Thermotoga maritima
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 21-Jul-2000
C:Accession: A72228
R:Nelson, K.E.; Clayton, R.A.; Gill, S.R.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hic
Garrett, M.W.; Stewart, A.M.; Cotton, M.D.; Pratt, M.S.; Phillips, C.A.; Richardson,
C.M.
Nature 399, 323-329, 1999
A:Title: Evidence for lateral gene transfer between Archaea and Bacteria from genome
A:Reference number: A72200; MUID:99287316
A:Accession: A72228
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-553 <ARN>
A:Cross-references: GB:AE001807; GB:AE000512; NID:g4982216; PIDN:AAD36719.1; PID:g498
A:Experimental source: strain MSB8
C:Genetics:
A:Gene: TM1652

Query Match 72.5%; Score 29; DB 2; Length 553;
Best Local Similarity 66.7%; Pred. No. 2e+02;
```

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||:|:|:|

Db 269 RLDMVKRV 277

RESULT 34

S65060

phytoene desaturase precursor - maize

C:Species: Zea mays (maize)

C>Date: 28-Oct-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000

C:Accession: S65060

R:Li, Z.H.; Matthews, P.D.; Burr, B.; Wurtzel, E.T.

Plant Mol. Biol. 30, 269-279, 1996

A:Title: Cloning and characterization of a maize cDNA encoding phytoene desaturase, an e

A:Reference number: S65060; MUID:96178866

A:Accession: S65060

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-571 <LI2>

A:Cross-references: EMBL:U37285; NID:g2707976; PIDN:AAC12846.1; PID:gl051180

C:Genetics:

A:Genome: nuclear

C:Keywords: chloroplast

Query Match 72.5%; Score 29; DB 2; Length 571;

Best Local Similarity 55.6%; Pred. No. 2e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||:|:|:|

Db 329 RLNSRIKKI 337

RESULT 35

H82872

hypothetical protein U0572 [imported] - Ureaplasma urealyticum

C:Species: Ureaplasma urealyticum

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: H82872

R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.

submitted to GenBank, February 2000

A:Description: The complete sequence of Ureaplasma urealyticum: Alternate views of a min

A:Reference number: A82870

A:Accession: H82872

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-578 <GLA>

A:Cross-references: GB:AE002156; GB:AF222894; NID:g6899580; PIDN:AAF30986.1; GSPDB:GN001

A:Experimental source: serovar 3; biovar 1

C:Genetics:

A:Gene: U0572

A:Genetic code: SGC3

Query Match 72.5%; Score 29; DB 2; Length 578;

Best Local Similarity 55.6%; Pred. No. 2.1e-02;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||:|:|:|

Db 98 RLNTVVKKI 106

RESULT 36

S29314

phytoene dehydrogenase (EC 1.3.1.20) - pepper

N:Alternate names: phytoene desaturase

C:Species: Capsicum annuum (pepper)

C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 21-Jul-2000

C:Accession: S29314

R:Hugueney, P.; Roemer, S.; Kuntz, M.; Camara, B.

Eur. J. Biochem. 209, 399-407, 1992

A:Title: Characterization and molecular cloning of a flavoprotein catalyzing the synt
A:Reference number: S29314; MUID:93011154
A:Accession: S29314
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-582 <HUG>
A:Cross-references: EMBL:X68058; NID:gl7950; PIDN:CAA48195.1; PID:gl7951
C:Keywords: chloroplast; oxidoreductase

Query Match 72.5%; Score 29; DB 2; Length 582;

Best Local Similarity 55.6%; Pred. No. 2.1e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||:|:|:|

Db 343 RLNSRIKKI 351

RESULT 37

A45381

phytoene dehydrogenase (EC 1.3.1.20) - tomato

N:Alternate names: phytoene desaturase

C:Species: Lycopersicon esculentum (tomato)

C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 21-Jul-2000

C:Accession: A45381; S42544; S21502

R:Pecker, I.; Chamovitz, D.; Linden, H.; Sandmann, G.; Hirschberg, J.

Proc. Natl. Acad. Sci. U.S.A. 89, 4962-4966, 1992

A:Title: A single polypeptide catalyzing the conversion of phytoene to zeta-carotene

A:Reference number: A45381; MUID:92279247

A:Accession: A45381

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-583 <PEC>

A:Cross-references: EMBL:X59948; NID:gl9286; PIDN:CAA42573.1; PID:gl9287

A:Experimental source: ripening fruit

A>Note: sequence extracted from NCBI backbone (NCBIP:104205)

R:Mann, V.; Pecker, I.; Hirschberg, J.

Plant Mol. Biol. 24, 429-434, 1994

A:Title: Cloning and characterization of the gene for phytoene desaturase (Pds) from

A:Reference number: S42544; MUID:94169297

A:Accession: S42544

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-583 <MAN>

A:Cross-references: EMBL:X71023; NID:g312822; PIDN:CAB59726.1; PID:g6138825

C:Genetics:

A:Introns: 75/3; 120/1; 149/3; 169/1; 221/1; 271/1; 310/3; 382/1; 416/2; 432/2; 447/3

C:Keywords: chloroplast; oxidoreductase

Query Match 72.5%; Score 29; DB 2; Length 583;

Best Local Similarity 55.6%; Pred. No. 2.1e+02;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||:|:|:|

Db 344 RLNSRIKKI 352

RESULT 38

T23338

hypothetical protein K05C4.5 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T23338

R:Harris, B.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19729

A:Accession: T23338

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-656 <WIL>

A:Cross-references: EMBL:281564; PIDN:CAB04569.1; GSPDB:GN00019; CESP:K05C4.5

A:Experimental source: clone K05C4

C:Genetics:
A:Gene: CESP:K05C4.5
A:Map position: 1
A:Introns: 8/2; 260/3; 428/1; 562/3; 621/1

Query Match 72.5%; Score 29; DB 2; Length 656;
Best Local Similarity 55.6%; Pred. No. 2.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
::| |||||
Db 366 KISPMVKKI 374

RESULT 39

ET0906
probable beta-glucosidase - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C:Accession: E70906
R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A:Authors: Sgates, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: E70906
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-691 <COL>
A:Cross-references: GB:297050; GB:AL123456; NID:g3256008; PIDN:CAB09737.1; PID:g2213523
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: bglS
C:Superfamily: beta-glucosidase

Query Match 72.5%; Score 29; DB 2; Length 691;
Best Local Similarity 55.6%; Pred. No. 2.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||| |||||
Db 282 RLSDMVRI 290

RESULT 40

S62963
hypothetical protein YNL041c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein N2675
C:Species: Saccharomyces cerevisiae
C:Date: 27-Apr-1996 #sequence_revision 03-May-1996 #text_change 05-Nov-1999
C:Accession: S62963
R:Duesterhoeft, A.; Floeth, M.; Fritz, C.; Heuss-Neitzel, D.; Hilbert, H.; Moestl, D.
submitted to the Protein Sequence Database, April 1996
A:Reference number: S62944
A:Accession: S62963
A:Molecule type: DNA
A:Residues: 1-839 <DUE>
A:Cross-references: EMBL:271317; NID:gl301893; PID:e239673; PID:gl301894; GSPDB:GN00014;
A:Experimental source: strain S288C
C:Genetics:
A:Gene: MIPS:YNL041c
A:Map position: 14L

Query Match 72.5%; Score 29; DB 2; Length 839;
Best Local Similarity 66.7%; Pred. No. 3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||||| |:
Db 227 RLSSSVKEI 235

RESULT 41

F96596
hypothetical protein TSA14.15 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: F96596
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzia
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: F96596
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1360 <STO>
A:Cross-references: GB:AE005173; NID:g4204269; PIDN:AAD10650.1; GSPDB:GN00141
C:Genetics:
A:Gene: TSA14.15
A:Map position: 1

Query Match 72.5%; Score 29; DB 2; Length 1360;
Best Local Similarity 85.7%; Pred. No. 4.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKK 8
:|||||
Db 711 ISSMVKK 717

RESULT 42

TI7484
hypothetical protein PCZA363.4 - Amycolatopsis orientalis
C:Species: Amycolatopsis orientalis
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 01-Dec-2000
C:Accession: TI7484
R:Van Wageningen, A.; Kirkpatrick, P.; Williams, D.; Harris, B.; Kershaw, J.; Lennard
Chem. Biol. 3, 155-162, 1998
A:Title: Sequencing and analysis of genes involved in the biosynthesis of a vancomycin
A:Reference number: Z18804
A:Accession: TI7484
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4077 <VAN>
A:Cross-references: EMBL:AJ223999; NID:e1251240; PID:e1251244; PIDN:CAAL1795.1
C:Superfamily: acyl carrier protein homology; acetate--CoA ligase homology
C:Keywords: carrier protein; phosphopantetheine; phosphoprotein
F:501-939/Domain: acetate--CoA ligase homology <ACL1>
F:955-1022/Domain: acyl carrier protein homology <ACP1>
F:2001-2441/Domain: acetate--CoA ligase homology <ACP1>
F:2457-2524/Domain: acyl carrier protein homology <ACP2>
F:3517-3956/Domain: acetate--CoA ligase homology <ACL3>
F:3972-4039/Domain: acyl carrier protein homology <ACP3>
F:987,2489,4004/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 72.5%; Score 29; DB 2; Length 4077;
Best Local Similarity 55.6%; Pred. No. 1.4e+03;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||:||||:
Db 4034 RLAAWVKEI 4042

RESULT 43

TI4914
dynein beta heavy chain - Tetrahymena thermophila

C:Species: Tetrahymena thermophila
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 21-Jan-2000
C:Accession: T14914
R:Lincoln, L.M.; Gibson, T.M.; Asai, D.J.; Forney, J.D.
submitted to the EMBL Data Library, June 1999
A:Description: A gene knockout reveals that dynein beta heavy chain is required in Tetra
A:Reference number: Z18264
A:Accession: T14914
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-4589 <LIN>
A:Cross-references: EMBL:AF072878; NID:g5209335; PIDN:AAC26117.1
C:Genetics:
A:Gene: DYH4
A:Genetic code: SGC5
A:Introns: 286/3; 4360/2; 4535/3
C:Superfamily: dynein heavy chain, ciliary
C:Keywords: P-loop

Query Match 72.5%; Score 29; DB 2; Length 4589;
Best Local Similarity 55.6%; Pred. No. 1.5e+03;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
II : : : :
Db 1761 RLENLIKV 1769

RESULT 44
G71376
hypothetical protein TP0010 - syphilis spirochete
C:Species: Treponema pallidum subsp. pallidum (syphilis spirochete)
C:Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 05-Nov-1999
C:Accession: G71376
R:Fraser, C.M.; Norris, S.J.; Weinstock, G.M.; White, O.; Sutton, G.G.; Dodson, R.; Gwin
rson, J.; Khalak, H.; Richardson, D.; Howell, J.K.; Chidambaram, M.; Utterback, T.; McDo
they, L.; Weidman, J.; Smith, H.O.; Venter, J.C.
Science 281, 375-388, 1998
A:Title: Complete genome sequence of Treponema pallidum, the syphilis spirochete.
A:Reference number: A71250; MUID:98332770
A:Accession: G71376
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-41 <COL>
A:Cross-references: GB:AE001186; GB:AE000520; NID:g3322263; PIDN:AAC65011.1; PID:g3322266
A:Experimental source: strain Nichols
C:Genetics:
A:Gene: TP0010

Query Match 70.0%; Score 28; DB 2; Length 41;
Best Local Similarity 55.6%; Pred. No. 27;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
: : : : :
Db 26 KLSEVVKI 34

RESULT 45
AH2565
hypothetical protein asr523 [imported] - Anabaena sp. (strain PCC 7120) plasmid pCC7120
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AH2565
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AH2565
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-63 <KUR>
A:Cross-references: GB:AP003604; PIDN:BAW77442.1; PID:g17134886; GSPDB:GN00183
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: asr523
A:Genome: plasmid

Query Match 70.0%; Score 28; DB 2; Length 63;
Best Local Similarity 66.7%; Pred. No. 41;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
II : : : :
Db 50 RLSSMVNQL 58

Search completed: November 6, 2002, 12:08:09
Job time : 14 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 23.7778 Seconds
(without alignments)
42.042 Million cell updates/sec

Title: US-09-905-083-32
Perfect score: 40
Sequence: 1 RLSSWVKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	9	22	AAE08237 Human Stratum Corn
2	40	100.0	225	22	AAE08502 Human Stratum Corn
3	40	100.0	253	16	AAE07888 Human Stratum Corn
4	40	100.0	253	17	AAE05383 Human amyloid prec
5	40	100.0	257	21	AAE21326 Human HSCCE. Homo
6	36	90.0	9	22	AAE08314 Human Stratum Corn
7	31	77.5	9	22	AAE08284 Human Stratum Corn
8	31	77.5	9	22	AAE08292 Human Stratum Corn
9	31	77.5	136	22	ABG23378 Novel human diagno
10	31	77.5	229	21	AAE19496 Arabidopsis thalia
11	31	77.5	229	21	AAE52872 Arabidopsis thalia

12	31	77.5	390	21	AAG19495 Arabidopsis thalia
13	31	77.5	390	21	AAG52871 Arabidopsis thalia
14	31	77.5	394	21	AAG19494 Arabidopsis thalia
15	31	77.5	394	21	AAG52870 Arabidopsis thalia
16	30	75.0	220	21	AAG24712 Plant SDF encoded
17	30	75.0	220	21	AAB25046 Arabidopsis thalia
18	30	75.0	220	21	AAG38343 Arabidopsis thalia
19	30	75.0	221	21	AAG38345 Arabidopsis thalia
20	30	75.0	236	21	AAB24711 Plant SDF encoded
21	30	75.0	236	21	AAB25045 Arabidopsis thalia
22	30	75.0	236	21	AAG38342 Arabidopsis thalia
23	30	75.0	237	21	AAG38344 Arabidopsis thalia
24	30	75.0	532	21	AAG30984 Arabidopsis thalia
25	30	75.0	533	21	AAG45409 Arabidopsis thalia
26	30	75.0	578	22	ABE58393 Drosophila melanog
27	30	75.0	647	16	AAW11978 Mycobacterium gall
28	29	72.5	74	21	AAG12180 Arabidopsis thalia
29	29	72.5	74	21	AAG43245 Arabidopsis thalia
30	29	72.5	364	22	ABG17953 Novel human diagno
31	29	72.5	582	17	AAW01103 N. benthamiana phy
32	29	72.5	582	17	AAW01104 N. benthamiana phy
33	29	72.5	582	17	AAW01105 N. tabacum phytoen
34	28	70.0	83	21	AAG33802 Arabidopsis thalia
35	28	70.0	87	22	AAO13253 Human polypeptide
36	28	70.0	100	22	AAO12825 Streptococcus pneu
37	28	70.0	106	19	AAW60941 Arabidopsis thalia
38	28	70.0	121	21	AAG33801 Human polypeptide
39	28	70.0	168	22	AAO2532 Arabidopsis thalia
40	28	70.0	175	15	AAE54097 Mouse PAP. Mus mu
41	28	70.0	231	21	AAG06740 Arabidopsis thalia
42	28	70.0	231	21	AAG10371 Arabidopsis thalia
43	28	70.0	231	21	AAG12990 Arabidopsis thalia
44	28	70.0	261	22	ABE69690 Drosophila melanog
45	28	70.0	328	22	AAU48010 Propionibacterium

ALIGNMENTS

RESULT 1
AAE08237
ID AAE08237 standard; peptide; 9 AA.
XX
AC AAE08237;
XX
DT 01-NOV-2001 (first entry)
XX
DE Human stratum corneum chymotrypsin enzyme peptide #2 (residues 123-131).
XX
KW Stratum corneum chymotrypsin enzyme; SCCE: cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antitense therapy; malignant hyperplasia.
XX
OS Homo sapiens.
XX
PN WO200159158-A1.
XX
PD 16-AUG-2001.
XX
PF 07-FEB-2001; 2001WO-US039977.
XX
PR 11-FEB-2000; 2000US-0502600.
XX
PA (UYAR-) UNIV ARKANSAS.
XX
PI O'brien TJ;
XX
DR WPI; 2001-514676/56.
XX
PT Diagnosing cancer comprises detecting stratum corneum chymotrypsin
enzyme -
XX
PS Claim 25; Page 102; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.

XX Sequence 9 AA;

Query Match 100.0%; Score 40; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLSSMVKKV 9
 Db 1 RLSSMVKKV 9

RESULT 2

ID AAB98502 standard; Protein; 225 AA.
 XX
 AC AAB98502;

DT 03-AUG-2001 (first entry)

DE Human Stratum Corneum Chymotryptic Enzyme, SCCE, catalytic domain.

KW Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;
 KW tumour antigen-derived gene 15; serine protease;
 KW Stratum Corneum Chymotryptic Enzyme; SCCE.

OS Homo sapiens.

PN WO200129056-A1.

PD 26-APR-2001.

PF 20-OCT-2000; 2000WO-US29095.

PR 20-OCT-1999; 99US-0421213.

PA (UYAR-) UNIV ARKANSAS.

PI O'Brien TJ, Tanimoto H;

DR WPI; 2001-381031/40.

PT Novel extracellular serine protease, termed tumor antigen-derived gene
 PT 15 protein overexpressed in carcinomas and DNA encoding it, for
 PT diagnosis, treatment, prevention of cancer, particularly breast,
 PT ovarian cancer

PS Example 10; Fig 1; 130pp; English.

XX The present invention relates to human tumour antigen-derived gene 15
 CC (TADG-15) protein and coding sequence (see AAH23601 and AAB98500).
 CC TADG-15 is an extracellular serine protease. It was found that TADG-15 is
 CC over-expressed in ovarian tumours. TADG-15 protein or its fragments of
 CC 9-20 residues that lack TADG-15 protease activity are useful for
 CC vaccinating an individual against TADG-15, having, suspected of having or
 CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a
 CC diagnostic or therapeutic target in cancer. The present sequence was used
 CC in a sequence homology alignment with the catalytic domain of TADG-15.

XX Sequence 225 AA;

Query Match 100.0%; Score 40; DB 22; Length 225;
 Best Local Similarity 100.0%; Pred. No. 0.98;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 RLSSMVKKV 9
 Db 95 RLSSMVKKV 103

RESULT 3

AAR67888
 ID AAR67888 standard; Protein; 253 AA.
 XX
 AC AAR67888;

DT 09-AUG-1995 (first entry)

DE Human stratum corneum chymotrophic recombinant enzyme (SCCE).

KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
 KW callosities; keratosis pilaris; ichthyoses; eczema.

OS Homo sapiens.

PN WO9500651-A.

PD 05-JAN-1995.

PF 20-JUN-1994; 94WO-IB00166.

PR 18-JUN-1993; 93DK-0000725.

PA (SYMB-) SYMBICOM AB.

PI Egelrud T, Hansson L;

DR WPI; 1995-052088/07.

DR N-PSDB; AAQ81203.

XX Nucleotide sequences encoding stratum corneum chymotryptic enzyme
 PT - and related vectors, transformed cells and polypeptides, and
 PT useful for treating skin disorders, e.g. acne or psoriasis, and
 PT for identification of specific inhibitors.

XX Disclosure; Page 97; 137pp; English.

XX The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
 CC and skin care products, especially to treat and prevent acne, eczema,
 CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
 CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
 CC produced recombinantly following mammal, insect, plant, or
 CC microorganism transformation with plasmid pS507.

SQ Sequence 253 AA;

Query Match 100.0%; Score 40; DB 16; Length 253;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLSSMVKKV 9

Db 123 RLSSMVKKV 131

RESULT 4

AAW05383
 ID AAW05383 standard; Protein; 253 AA.
 XX
 AC AAW05383;

DT 31-DEC-1996 (first entry)

DE Human amyloid precursor protein protease.

KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;

KW therapy.
 XX Homo sapiens.
 XX WO9631122-A1.
 PN 10-OCT-1996.
 XX
 XX
 XX 02-APR-1996; 96WO-US04294.
 XX
 PR 04-APR-1995; 95US-0416257.
 XX
 XX (ELIL) LILLY & CO ELI.
 XX
 XX Dixon EP, Johnstone EM, Little SP;
 PI
 XX WPI; 1996-464694/46.
 DR N-PSDB; AAT39783.
 XX
 XX New isolated human amyloid precursor protein protease - used to
 PT develop prods. for the treatment or diagnosis of associated
 PT conditions, esp. Alzheimer's disease
 PT
 XX
 PS Claim 1; Page 44-45; 55pp; English.
 XX
 CC Human amyloid precursor protein protease (AAW05383) is involved in
 CC the processing or clearance of amyloid precursor protein to form
 CC beta-amyloid peptide. Its amino acid sequence was deduced from
 CC a cDNA clone (AAT39783) obtd. from a human lung library. Recombinant
 CC protease can be produced in transformed or transfected prokaryotic
 CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
 CC used to develop products for the design and testing of cpds. useful
 CC for treating or preventing conditions associated with beta-amyloid
 CC peptide, esp. Alzheimer's disease.
 XX
 SQ Sequence 253 AA;
 Query Match 100.0%; Score 40; DB 17; Length 253;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RLSSMVKKV 9
 Db 123 RLSSMVKKV 131
 |||||
 RESULT 5
 AAB21326
 ID AAB21326 standard; Protein; 257 AA.
 XX
 AC AAB21326;
 XX
 XX 02-FEB-2001 (first entry)
 XX
 DE Human HSCCE.
 XX
 KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCCE;
 KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;
 KW serine protease; cytostatic; cancer; prostrate cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200053776-A2.
 XX
 PD 14-SEP-2000.
 XX
 XX 09-MAR-2000; 2000WO-CA00258.
 XX
 PR 11-MAR-1999; 99US-0124260.
 PR 01-APR-1999; 99US-0127386.
 PR 21-JUL-1999; 99US-0144919.
 XX
 XX (MOUN) MOUNT SINAI HOSPITAL.

XX Yousef GM, Diamandis EP;
 PI
 XX WPI; 2000-587440/55.
 XX
 XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
 PT protein mediated disorders, especially cancer. -
 XX
 XX Example 4; Fig 17; 184pp; English.
 XX
 CC The present sequence is human stratum corneum chymotryptic enzyme
 CC (HSCCE), a member of the kallikrein multi-gene family. Kallikreins and
 CC kallikrein-like proteins are a subgroup of the serine protease enzyme
 CC family. They catalyse the selective cleavage of specific polypeptide
 CC precursors to release peptides with potent biological activity. Nucleic
 CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
 CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
 CC treatment, monitoring and diagnosis of cancers, especially prostate
 CC cancer. They can also be used to identify a substance that can associate
 CC with or mediate the biological activity of the proteins. Antibodies can
 CC be used to treat conditions mediated by the kallikrein-like proteins.
 XX
 SQ Sequence 257 AA;
 Query Match 100.0%; Score 40; DB 21; Length 257;
 Best Local Similarity 100.0%; Pred. No. 1.1;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RLSSMVKKV 9
 Db 128 RLSSMVKKV 136
 |||||
 RESULT 6
 AAE08314
 ID AAE08314 standard; peptide; 9 AA.
 XX
 AC AAE08314;
 XX
 DT 01-NOV-2001 (first entry)
 XX
 DE Human stratum corneum chymotrypsin enzyme peptide #79 (residues 122-130).
 XX
 KW Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
 KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
 KW antisense therapy; malignant hyperplasia.
 XX
 OS Homo sapiens.
 XX
 PN WO200159158-A1.
 XX
 PD 16-AUG-2001.
 XX
 XX 07-FEB-2001; 2001WO-US03977.
 PF
 XX 11-FEB-2000; 2000US-0502600.
 PR
 XX (UYAR-) UNIV ARKANSAS.
 PA
 XX O'brien TJ;
 PI
 XX WPI; 2001-514676/56.
 DR
 XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
 PT enzyme -
 PT
 XX Disclosure; Page 120; 127pp; English.
 PS
 CC The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,

CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.
XX
SQ Sequence 9 AA;

Query Match 90.0%; Score 36; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMKVKK 8
 |||||||
DB 2 RLSSMKVKK 9

RESULT 7

AAE08284
ID AAE08284 standard; peptide; 9 AA.

XX AC AAE08284;

DT 01-NOV-2001 (first entry)

XX Human stratum corneum chymotrypsin enzyme peptide #49 (residues 125-133).

XX Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.

XX Homo sapiens.

XX WO200159158-A1.

PD 16-AUG-2001.

XX 07-FEB-2001; 2001WO-US03977.

XX 11-FEB-2000; 2000US-0502600.

PA (UYAR-) UNIV ARKANSAS.

XX O'Brien TJ;

XX WPI; 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -

PS Disclosure; Page 113; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.

XX Sequence 9 AA;

Query Match 77.5%; Score 31; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SSMVKVKK 9
 |||||||
DB 1 SSMVKVKK 7

RESULT 8

AAE08292

ID AAE08292 standard; peptide; 9 AA.

XX AC AAE08292;

DT 01-NOV-2001 (first entry)

XX Human stratum corneum chymotrypsin enzyme peptide #57 (residues 125-133).

XX Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.

XX Homo sapiens.

XX WO200159158-A1.

PD 16-AUG-2001.

XX 07-FEB-2001; 2001WO-US03977.

XX 11-FEB-2000; 2000US-0502600.

PA (UYAR-) UNIV ARKANSAS.

XX O'Brien TJ;

XX WPI; 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -

PS Disclosure; Page 115; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.

XX Sequence 9 AA;

Query Match 77.5%; Score 31; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SSMVKVKK 9
 |||||||
DB 1 SSMVKVKK 7

RESULT 9

ABG23378
ID ABG23378 standard; Protein; 136 AA.

XX AC ABG23378;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #23369.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.

PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.5%; Score 31; DB 21; Length 229;
Best Local Similarity 55.6%; Pred No. 84;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMKKV 9
| | | | |
Db 69 RFSSMIRKI 77

RESULT 11
AAG52872
ID AAG52872 standard; Protein; 229 AA.
XX
AC AAG52872;

DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 67256.

XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX
OS Arabidopsis thaliana.

XX
PN EP1033405-A2.

XX
PD 06-SEP-2000.

XX
PF 25-FEB-2000; 2000EP-0301439.

XX
PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123546.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136592.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142055.
PR 08-JUL-1999; 99US-0142390.
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 28-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.

PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159337.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 77.5%; Score 31; DB 21; Length 229;

Best Local Similarity 55.6%; Pred. No. 84;

Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSWKKV 9

DB 69 RFSSMIRKI 77

RESULT 12

AAG19495
 ID AAG19495 standard; Protein; 390 AA.

XX AC AAG19495;

XX XX

DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21320.

XX KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.

XX OS Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 25-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139452.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139753.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.

PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.5%; Score 31; DB 21; Length 390;
Best Local Similarity 55.6%; Pred. NO. 1.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMKVKV 9
| | | | | | | | | |
Db 230 RFSSMKIRKI 238

RESULT 13
AAG52871
ID AAG52871 standard; Protein; 390 AA.
XX AC AAG52871;
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 67255.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.

PR	06-MAY-1999;	990S-0132486.	PR	22-JUL-1999;	990S-0145089.
PR	06-MAY-1999;	990S-0132487.	PR	22-JUL-1999;	990S-0145192.
PR	07-MAY-1999;	990S-0132863.	PR	23-JUL-1999;	990S-0145149.
PR	11-MAY-1999;	990S-0134256.	PR	23-JUL-1999;	990S-0145218.
PR	14-MAY-1999;	990S-0134218.	PR	23-JUL-1999;	990S-0145224.
PR	14-MAY-1999;	990S-0134219.	PR	26-JUL-1999;	990S-0145276.
PR	14-MAY-1999;	990S-0134221.	PR	27-JUL-1999;	990S-0145913.
PR	14-MAY-1999;	990S-0134370.	PR	27-JUL-1999;	990S-0145918.
PR	18-MAY-1999;	990S-0134768.	PR	27-JUL-1999;	990S-0145919.
PR	19-MAY-1999;	990S-0134941.	PR	28-JUL-1999;	990S-0145951.
PR	20-MAY-1999;	990S-0135124.	PR	02-AUG-1999;	990S-0146386.
PR	21-MAY-1999;	990S-0135353.	PR	02-AUG-1999;	990S-0146388.
PR	24-MAY-1999;	990S-0135629.	PR	02-AUG-1999;	990S-0146389.
PR	25-MAY-1999;	990S-0136021.	PR	03-AUG-1999;	990S-0147038.
PR	27-MAY-1999;	990S-0136392.	PR	04-AUG-1999;	990S-0147204.
PR	28-MAY-1999;	990S-0136782.	PR	04-AUG-1999;	990S-0147302.
PR	01-JUN-1999;	990S-0137222.	PR	05-AUG-1999;	990S-0147192.
PR	03-JUN-1999;	990S-0137528.	PR	05-AUG-1999;	990S-0147260.
PR	04-JUN-1999;	990S-0137502.	PR	06-AUG-1999;	990S-0147300.
PR	07-JUN-1999;	990S-0137724.	PR	06-AUG-1999;	990S-0147416.
PR	08-JUN-1999;	990S-0138094.	PR	09-AUG-1999;	990S-0147493.
PR	10-JUN-1999;	990S-0138540.	PR	09-AUG-1999;	990S-0147935.
PR	10-JUN-1999;	990S-0138847.	PR	10-AUG-1999;	990S-0148171.
PR	14-JUN-1999;	990S-0139119.	PR	11-AUG-1999;	990S-0148319.
PR	16-JUN-1999;	990S-0139452.	PR	12-AUG-1999;	990S-0148341.
PR	16-JUN-1999;	990S-0139453.	PR	13-AUG-1999;	990S-0148565.
PR	17-JUN-1999;	990S-0139492.	PR	13-AUG-1999;	990S-0148684.
PR	18-JUN-1999;	990S-0139454.	PR	16-AUG-1999;	990S-0149368.
PR	18-JUN-1999;	990S-0139455.	PR	17-AUG-1999;	990S-0149175.
PR	18-JUN-1999;	990S-0139456.	PR	18-AUG-1999;	990S-0149426.
PR	18-JUN-1999;	990S-0139457.	PR	20-AUG-1999;	990S-0149722.
PR	18-JUN-1999;	990S-0139458.	PR	20-AUG-1999;	990S-0149723.
PR	18-JUN-1999;	990S-0139459.	PR	20-AUG-1999;	990S-0149929.
PR	18-JUN-1999;	990S-0139460.	PR	23-AUG-1999;	990S-0149902.
PR	18-JUN-1999;	990S-0139461.	PR	23-AUG-1999;	990S-0149930.
PR	18-JUN-1999;	990S-0139462.	PR	25-AUG-1999;	990S-0150566.
PR	18-JUN-1999;	990S-0139463.	PR	26-AUG-1999;	990S-0150884.
PR	18-JUN-1999;	990S-0139750.	PR	27-AUG-1999;	990S-0151065.
PR	18-JUN-1999;	990S-0139763.	PR	27-AUG-1999;	990S-0151066.
PR	21-JUN-1999;	990S-0139817.	PR	30-AUG-1999;	990S-0151080.
PR	22-JUN-1999;	990S-0139899.	PR	30-AUG-1999;	990S-0151303.
PR	23-JUN-1999;	990S-0140353.	PR	31-AUG-1999;	990S-0151438.
PR	23-JUN-1999;	990S-0140354.	PR	01-SEP-1999;	990S-0151930.
PR	24-JUN-1999;	990S-0140695.	PR	07-SEP-1999;	990S-0152363.
PR	28-JUN-1999;	990S-0140823.	PR	10-SEP-1999;	990S-0153070.
PR	29-JUN-1999;	990S-0140991.	PR	13-SEP-1999;	990S-0153758.
PR	30-JUN-1999;	990S-0141287.	PR	15-SEP-1999;	990S-0154018.
PR	01-JUL-1999;	990S-0141842.	PR	16-SEP-1999;	990S-0154039.
PR	01-JUL-1999;	990S-0142154.	PR	20-SEP-1999;	990S-0154779.
PR	02-JUL-1999;	990S-0142055.	PR	22-SEP-1999;	990S-0155139.
PR	06-JUL-1999;	990S-0142390.	PR	23-SEP-1999;	990S-0155486.
PR	08-JUL-1999;	990S-0142803.	PR	24-SEP-1999;	990S-0155659.
PR	09-JUL-1999;	990S-0142920.			

```
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 23-OCT-1999; 99US-0160980.
PR 24-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0160989.
PR 26-OCT-1999; 99US-0161404.
PR 27-OCT-1999; 99US-0161405.
PR 28-OCT-1999; 99US-0161406.
PR 29-OCT-1999; 99US-0161359.
PR 30-OCT-1999; 99US-0161360.
PR 31-OCT-1999; 99US-0161361.
PR 01-NOV-1999; 99US-0161920.
PR 02-NOV-1999; 99US-0161921.
PR 03-NOV-1999; 99US-0161992.
PR 04-NOV-1999; 99US-0161993.
PR 05-NOV-1999; 99US-0162142.

Query Match      77.5%; Score 31; DB 21; Length 390;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSVKKV 9
   | | | | |
DB 230 RFSSMIRKI 238

RESULT 14
AAG19494
ID AAG19494 standard; Protein; 394 AA.
XX AC AAG19494;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 21319.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132407.
PR 06-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132487.
PR 14-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.

14-MAY-1999; 99US-0134370.
18-MAY-1999; 99US-0134768.
19-MAY-1999; 99US-0134941.
20-MAY-1999; 99US-0135124.
21-MAY-1999; 99US-0135353.
24-MAY-1999; 99US-0135629.
25-MAY-1999; 99US-0136021.
27-MAY-1999; 99US-0136392.
28-MAY-1999; 99US-0136782.
01-JUN-1999; 99US-0137222.
03-JUN-1999; 99US-0137528.
04-JUN-1999; 99US-0137502.
07-JUN-1999; 99US-0137724.
08-JUN-1999; 99US-0138094.
10-JUN-1999; 99US-0138540.
10-JUN-1999; 99US-0138847.
14-JUN-1999; 99US-0139119.
16-JUN-1999; 99US-0139452.
16-JUN-1999; 99US-0139453.
17-JUN-1999; 99US-0139492.
18-JUN-1999; 99US-0139454.
18-JUN-1999; 99US-0139455.
18-JUN-1999; 99US-0139456.
18-JUN-1999; 99US-0139457.
18-JUN-1999; 99US-0139458.
18-JUN-1999; 99US-0139459.
18-JUN-1999; 99US-0139460.
18-JUN-1999; 99US-0139461.
18-JUN-1999; 99US-0139462.
18-JUN-1999; 99US-0139463.
18-JUN-1999; 99US-0139750.
18-JUN-1999; 99US-0139763.
21-JUN-1999; 99US-0139817.
22-JUN-1999; 99US-0139899.
23-JUN-1999; 99US-0140353.
23-JUN-1999; 99US-0140354.
24-JUN-1999; 99US-0140695.
28-JUN-1999; 99US-0140823.
29-JUN-1999; 99US-0140991.
30-JUN-1999; 99US-0141287.
01-JUL-1999; 99US-0141842.
01-JUL-1999; 99US-0142154.
02-JUL-1999; 99US-0142055.
06-JUL-1999; 99US-0142390.
08-JUL-1999; 99US-0142803.
09-JUL-1999; 99US-0142920.
12-JUL-1999; 99US-0142977.
13-JUL-1999; 99US-0143542.
14-JUL-1999; 99US-0143624.
15-JUL-1999; 99US-0144005.
16-JUL-1999; 99US-0144085.
16-JUL-1999; 99US-0144086.
19-JUL-1999; 99US-0144325.
19-JUL-1999; 99US-0144331.
19-JUL-1999; 99US-0144332.
19-JUL-1999; 99US-0144333.
19-JUL-1999; 99US-0144334.
19-JUL-1999; 99US-0144335.
20-JUL-1999; 99US-0144352.
20-JUL-1999; 99US-0144632.
20-JUL-1999; 99US-0144884.
21-JUL-1999; 99US-0144814.
21-JUL-1999; 99US-0145086.
21-JUL-1999; 99US-0145088.
22-JUL-1999; 99US-0145085.
22-JUL-1999; 99US-0145087.
22-JUL-1999; 99US-0145089.
22-JUL-1999; 99US-0145192.
23-JUL-1999; 99US-0145145.
23-JUL-1999; 99US-0145218.
23-JUL-1999; 99US-0145224.
26-JUL-1999; 99US-0145276.
27-JUL-1999; 99US-0145913.
```

```
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157153.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.

PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.5%; Score 31; DB 21; Length 394;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLSSMKV 9
Db 234 RFSSMKIRKI 242

RESULT 15
AAG52870
ID AAG52870 standard; Protein; 394 AA.
XX AC AAG52870;
XX DT 18-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 67254.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
```

PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141682.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144884.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 05-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.
PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.

PR 29-OCT-1999; 99US-0162142.

Query Match 77.5%; Score 31; DB 21; Length 394;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
I : : : : :
Db 234 RFSSMIRKI 242

RESULT 16

AAB24712
ID AAB24712 standard; Peptide; 220 AA.

XX

AC AAB24712;

XX 27-NOV-2000 (first entry)

XX Plant SDF encoded polypeptide sequence SEQ List 1 NO:152.

XX Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;
KW SDF; genetic mapping; identification; promoter; structural gene; UTR;
KW untranslated region; expression control.

XX OS Plant.

XX WO2000040695-A2.

XX 13-JUL-2000.

XX 07-JAN-2000; 2000WO-US00466.

XX 08-JAN-1999; 99US-0115293.

XX (CERE-) CERES INC.

XX Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L;

XX WPI; 2000-465970/40.

XX New corn plant and Arabidopsis thaliana sequence-determined DNA
PT fragments, useful for expressing gene products and for controlling
PT expression of a target gene -

XX Claim 14; Page 395-396; 673pp; English.

XX The present invention describes polynucleotides, such as complete cDNA
CC sequences and/or sequences of genomic DNA encompassing complete genes,
CC portions of genes, and/or intergenic regions, collectively referred to
CC as sequence-determined DNA fragments (SDFs), from corn plants and
CC Arabidopsis thaliana. The SDFs are promoters, structural genes,
CC untranslated regions (UTRs), or 3' termination sequences. They can be
CC used for expressing a gene product and controlling expression of a
CC target gene, either as a promoter, a structural gene, an UTR or as a
CC 3' termination sequence. They are also useful as tools for genetic
CC mapping, and identification of a particular individual plant or for
CC clustering a group of plants with a common trait. AAA78433 to AAA78630
CC and AAB24605 to AAB25099 represent the specifically claimed
CC polynucleotide sequences and polypeptides encoded by them given in the
CC present invention.

XX Sequence 220 AA;

Query Match 75.0%; Score 30; DB 21; Length 220;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
I : : : : :
Db 132 RLSSSMKKI 140

RESULT 18

AAG38343
ID AAG38343 standard; Protein; 220 AA.

XX AAG38343;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 47290.

RESULT 17

AAB25046
ID AAB25046 standard; Peptide; 220 AA.

XX AAB25046;

XX 27-NOV-2000 (first entry)

XX Plant SDF encoded polypeptide sequence SEQ List 2 NO:90.

XX Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;
KW SDF; genetic mapping; identification; promoter; structural gene; UTR;
KW untranslated region; expression control.

XX OS Plant.

XX WO2000040695-A2.

XX 13-JUL-2000.

XX 07-JAN-2000; 2000WO-US00466.

XX 08-JAN-1999; 99US-0115293.

XX (CERE-) CERES INC.

XX Alexandrov N, Brover V, Chen X, Subramanian G, Troukhan ME;
PI Zheng L;

XX WPI; 2000-465970/40.

XX New corn plant and Arabidopsis thaliana sequence-determined DNA
PT fragments, useful for expressing gene products and for controlling
PT expression of a target gene -

XX Claim 14; Page 634; 673pp; English.

XX The present invention describes polynucleotides, such as complete cDNA
CC sequences and/or sequences of genomic DNA encompassing complete genes,
CC portions of genes, and/or intergenic regions, collectively referred to
CC as sequence-determined DNA fragments (SDFs), from corn plants and
CC Arabidopsis thaliana. The SDFs are promoters, structural genes,
CC untranslated regions (UTRs), or 3' termination sequences. They can be
CC used for expressing a gene product and controlling expression of a
CC target gene, either as a promoter, a structural gene, an UTR or as a
CC 3' termination sequence. They are also useful as tools for genetic
CC mapping, and identification of a particular individual plant or for
CC clustering a group of plants with a common trait. AAA78433 to AAA78630
CC and AAB24605 to AAB25099 represent the specifically claimed
CC polynucleotide sequences and polypeptides encoded by them given in the
CC present invention.

XX Sequence 220 AA;

Query Match 75.0%; Score 30; DB 21; Length 220;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
I : : : : :
Db 132 RLSSSMKKI 140

xx Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
xx Arabidopsis thaliana.
xx EP1033405-A2.
xx 06-SEP-2000.
xx
xx 25-FEB-2000; 2000EP-0301439.
xx
xx 25-FEB-1999; 99US-0121825.
xx 05-MAR-1999; 99US-0123180.
xx 09-MAR-1999; 99US-0123548.
xx 23-MAR-1999; 99US-0125788.
xx 25-MAR-1999; 99US-0126264.
xx 29-MAR-1999; 99US-0126785.
xx 01-APR-1999; 99US-0127462.
xx 06-APR-1999; 99US-0128234.
xx 08-APR-1999; 99US-0128714.
xx 16-APR-1999; 99US-0129845.
xx 19-APR-1999; 99US-0130077.
xx 21-APR-1999; 99US-0130449.
xx 23-APR-1999; 99US-0130510.
xx 28-APR-1999; 99US-0130891.
xx 30-APR-1999; 99US-0131449.
xx 30-APR-1999; 99US-0132048.
xx 04-MAY-1999; 99US-0132407.
xx 05-MAY-1999; 99US-0132484.
xx 06-MAY-1999; 99US-0132485.
xx 06-MAY-1999; 99US-0132486.
xx 07-MAY-1999; 99US-0132487.
xx 11-MAY-1999; 99US-0132863.
xx 14-MAY-1999; 99US-0134256.
xx 14-MAY-1999; 99US-0134218.
xx 14-MAY-1999; 99US-0134219.
xx 14-MAY-1999; 99US-0134221.
xx 14-MAY-1999; 99US-0134370.
xx 18-MAY-1999; 99US-0134768.
xx 19-MAY-1999; 99US-0134941.
xx 20-MAY-1999; 99US-0135124.
xx 21-MAY-1999; 99US-0135353.
xx 24-MAY-1999; 99US-0135629.
xx 25-MAY-1999; 99US-0136021.
xx 27-MAY-1999; 99US-0136392.
xx 28-MAY-1999; 99US-0136782.
xx 01-JUN-1999; 99US-0137222.
xx 03-JUN-1999; 99US-0137528.
xx 04-JUN-1999; 99US-0137502.
xx 07-JUN-1999; 99US-0137724.
xx 08-JUN-1999; 99US-0138094.
xx 10-JUN-1999; 99US-0138540.
xx 10-JUN-1999; 99US-0138847.
xx 14-JUN-1999; 99US-0139119.
xx 16-JUN-1999; 99US-0139452.
xx 17-JUN-1999; 99US-0139453.
xx 18-JUN-1999; 99US-0139492.
xx 18-JUN-1999; 99US-0139454.
xx 18-JUN-1999; 99US-0139455.
xx 18-JUN-1999; 99US-0139456.
xx 18-JUN-1999; 99US-0139457.
xx 18-JUN-1999; 99US-0139458.
xx 18-JUN-1999; 99US-0139459.
xx 18-JUN-1999; 99US-0139460.
xx 18-JUN-1999; 99US-0139461.
xx 18-JUN-1999; 99US-0139462.
xx 18-JUN-1999; 99US-0139463.
xx 18-JUN-1999; 99US-0139750.
xx 21-JUN-1999; 99US-0139763.
xx 22-JUN-1999; 99US-0139817.
xx 22-JUN-1999; 99US-0139899.
xx 23-JUN-1999;
xx 23-JUN-1999;
xx 24-JUN-1999;
xx 28-JUN-1999;
xx 29-JUN-1999;
xx 30-JUN-1999;
xx 01-JUL-1999;
xx 01-JUL-1999;
xx 02-JUL-1999;
xx 06-JUL-1999;
xx 08-JUL-1999;
xx 09-JUL-1999;
xx 12-JUL-1999;
xx 13-JUL-1999;
xx 14-JUL-1999;
xx 15-JUL-1999;
xx 16-JUL-1999;
xx 16-JUL-1999;
xx 19-JUL-1999;
xx 19-JUL-1999;
xx 19-JUL-1999;
xx 19-JUL-1999;
xx 19-JUL-1999;
xx 20-JUL-1999;
xx 20-JUL-1999;
xx 21-JUL-1999;
xx 21-JUL-1999;
xx 22-JUL-1999;
xx 22-JUL-1999;
xx 22-JUL-1999;
xx 22-JUL-1999;
xx 23-JUL-1999;
xx 23-JUL-1999;
xx 26-JUL-1999;
xx 27-JUL-1999;
xx 27-JUL-1999;
xx 27-JUL-1999;
xx 27-JUL-1999;
xx 28-JUL-1999;
xx 28-JUL-1999;
xx 02-AUG-1999;
xx 02-AUG-1999;
xx 02-AUG-1999;
xx 03-AUG-1999;
xx 04-AUG-1999;
xx 04-AUG-1999;
xx 05-AUG-1999;
xx 06-AUG-1999;
xx 06-AUG-1999;
xx 09-AUG-1999;
xx 09-AUG-1999;
xx 10-AUG-1999;
xx 11-AUG-1999;
xx 12-AUG-1999;
xx 13-AUG-1999;
xx 13-AUG-1999;
xx 16-AUG-1999;
xx 17-AUG-1999;
xx 18-AUG-1999;
xx 20-AUG-1999;
xx 20-AUG-1999;
xx 23-AUG-1999;
xx 23-AUG-1999;
xx 25-AUG-1999;
xx 26-AUG-1999;
xx 27-AUG-1999;
xx 27-AUG-1999;
xx 30-AUG-1999;
99US-0140353.
99US-0140354.
99US-0140695.
99US-0140823.
99US-0140991.
99US-0141287.
99US-0141842.
99US-0142154.
99US-0142055.
99US-0142390.
99US-0142803.
99US-0142920.
99US-0142977.
99US-0143542.
99US-0143624.
99US-0144005.
99US-0144085.
99US-0144086.
99US-0144325.
99US-0144331.
99US-0144332.
99US-0144333.
99US-0144334.
99US-0144335.
99US-0144352.
99US-0144632.
99US-0144884.
99US-0144814.
99US-0145086.
99US-0145088.
99US-0145085.
99US-0145087.
99US-0145089.
99US-0145192.
99US-0145145.
99US-0145218.
99US-0145224.
99US-0145276.
99US-0145913.
99US-0145918.
99US-0145919.
99US-0145951.
99US-0146386.
99US-0146388.
99US-0146389.
99US-0147038.
99US-0147204.
99US-0147302.
99US-0147192.
99US-0147260.
99US-0147303.
99US-0147416.
99US-0147493.
99US-0147935.
99US-0148171.
99US-0148319.
99US-0148341.
99US-0148565.
99US-0148684.
99US-0149368.
99US-0149175.
99US-0149426.
99US-0149722.
99US-0149723.
99US-0149929.
99US-0149902.
99US-0149930.
99US-0150566.
99US-0150884.
99US-0151065.
99US-0151066.
99US-0151080.
99US-0151303.


```
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.

PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 75.0%; Score 30; DB 21; Length 221;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
Db 132 RLSSSMKKI 140

RESULT 20
AAB24711
ID AAB24711 standard; Peptide; 236 AA.
XX
AC AAB24711;
XX
DT 27-NOV-2000 (first entry)
XX
DE Plant SDF encoded polypeptide sequence SEQ List 1 NO:151.
KW Plant; corn; Arabidopsis thaliana; sequence-determined DNA fragment;
KW SDF; genetic mapping; identification; promoter; structural gene; UTR;
KW untranslated region; expression control.
XX Plant.
OS
XX WO200040695-A2.
XX PN
XX 13-JUL-2000.
XX PD
XX 07-JAN-2000; 2000WO-US00466.
XX PF
XX 08-JAN-1999; 99US-0115293.
XX PR
```


PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.

```
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 75.0%; Score 30; DB 21; Length 236;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMKVKV 9
DB 148 RLSSMKKI 156

RESULT 23
AAG38344
ID AAG38344 standard; Protein; 237 AA.
XX
AC AAG38344;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 47292.
XX
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123546.
PR 23-MAR-1999; 99US-0125788.
PR 29-MAR-1999; 99US-0126264.
PR 01-APR-1999; 99US-0126785.
PR 06-APR-1999; 99US-0127462.
PR 08-APR-1999; 99US-0128234.
PR 16-APR-1999; 99US-0128714.
PR 19-APR-1999; 99US-0129845.
PR 21-APR-1999; 99US-0130077.
PR 23-APR-1999; 99US-0130449.
PR 28-APR-1999; 99US-0130510.
PR 30-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145513.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
```

```
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147036.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.
PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0158293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 21-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.

PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161992.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

Query Match          75.0%; Score 30; DB 21; Length 237;
Best Local Similarity 66.7%; Pred No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
Db 148 RLSSSMKKI 156

RESULT 24
AAG30984
ID AAG30984 standard; Protein; 532 AA.
XX
AC AAG30984;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 37137.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
PR 05-MAR-1999; 990S-0123180.
PR 09-MAR-1999; 990S-0123548.
PR 23-MAR-1999; 990S-0125788.
PR 25-MAR-1999; 990S-0126264.
PR 29-MAR-1999; 990S-0126785.
PR 01-APR-1999; 990S-0127462.
PR 06-APR-1999; 990S-0128234.
PR 08-APR-1999; 990S-0128714.
PR 16-APR-1999; 990S-0129845.
PR 19-APR-1999; 990S-0130077.
PR 21-APR-1999; 990S-0130449.
PR 23-APR-1999; 990S-0130510.
PR 23-APR-1999; 990S-0130891.
PR 28-APR-1999; 990S-0131449.
PR 30-APR-1999; 990S-0132048.
PR 30-APR-1999; 990S-0132407.
PR 04-MAY-1999; 990S-0132484.
PR 05-MAY-1999; 990S-0132485.
PR 06-MAY-1999; 990S-0132486.
PR 06-MAY-1999; 990S-0132487.
PR 07-MAY-1999; 990S-0132863.
PR 11-MAY-1999; 990S-0134256.
PR 14-MAY-1999; 990S-0134218.
PR 14-MAY-1999; 990S-0134219.
PR 14-MAY-1999; 990S-0134221.
PR 14-MAY-1999; 990S-0134370.
PR 18-MAY-1999; 990S-0134768.
PR 19-MAY-1999; 990S-0134941.
PR 20-MAY-1999; 990S-0135124.
PR 21-MAY-1999; 990S-0135353.
PR 24-MAY-1999; 990S-0135629.
PR 25-MAY-1999; 990S-0136021.
PR 27-MAY-1999; 990S-0136392.
PR 28-MAY-1999; 990S-0136782.
```

PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 29-JUN-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.


```
Query Match          75.0%; Score 30; DB 21; Length 532;
Best Local Similarity 87.5%; Pred. No. 3.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLSSMYKK 8
DB 7 RLSSMYKK 14

RESULT 25
AAG45409
ID AAG45409 standard; Protein; 533 AA.
XX
AC AAG45409;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 57006.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
```

PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159337.
 PR 18-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 22-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 25-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 75.0%; Score 30; DB 21; Length 533;
 Best Local Similarity 87.5%; Pred. No. 3.5e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLSSMVKK 8
 |||||
 Db 7 RLSSSVKK 14

RESULT 26

ABB58393
 ID ABB58393 standard; Protein; 578 AA.

XX AC ABB58393;

XX DT 26-MAR-2002 (first entry)

XX DE Drosophila melanogaster polypeptide SEQ ID NO 1971.

XX KW Drosophila; developmental biology; cell signalling; insecticide; pharmaceutical.

XX OS Drosophila melanogaster.

XX PN WO200171042-A2.

XX PD 27-SEP-2001.

XX PF 23-MAR-2001; 2001WO-US09231.

XX PR 23-MAR-2000; 2000US-191637P.

XX PR 11-JUL-2000; 2000US-0614150.

XX PA (PEKE) PE CORP NY.

XX PI Venter JC, Adams M, Li PWD, Myers EW;

XX DR WPI; 2001-656860/75.

XX DR N-PSDB; ABL02496.

XX PT New isolated nucleic acid detection reagent for detecting 1000 or more

XX PT genes from Drosophila and for elucidating cell signalling and cell-cell

XX PT interactions -

XX PS Disclosure; SEQ ID NO 1971; 21pp + Sequence Listing; English.

XX CC The invention relates to an isolated nucleic acid detection reagent

XX CC capable of detecting 1000 or more genes from Drosophila. The invention is

XX CC useful in developmental biology and in elucidating cell signalling and

XX CC cell-cell interactions in higher eukaryotes for the development of

XX CC insecticides, therapeutics and pharmaceutical drugs. The invention

XX CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA

XX CC sequences (ABL01840-ABL16175) and the encoded proteins

XX CC (ABB57737-ABB72072).

XX CC The sequence data for this patent did not form part of the printed

XX CC specification, but was obtained in electronic format directly from WIPO

XX CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 578 AA;

Query Match 75.0%; Score 30; DB 22; Length 578;

Best Local Similarity 75.0%; Pred. No. 3.8e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLSSMVKK 8

Db 249 RLNNMVKK 256

||::|||

RESULT 27

AAW11978
 ID AAW11978 standard; Protein; 647 AA.

XX AC AAW11978;

XX DT 23-APR-1997 (first entry)

XX DE Mycobacterium gallisepticum pMGAL.2 adhesin.

XX XX Adhesin; pMGA; mycoplasma; diagnosis; vaccine; vector;

[illegible]

PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 72.5%; Score 29; DB 21; Length 74;
Best Local Similarity 77.8%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
|||
Db 36 RLRSCKVKV 44

RESULT 29
AAG43245
ID AAG43245 standard; Protein; 74 AA.

XX AAG43245;

DT 18-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 54030.

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

OS Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.

PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 72.5%; Score 29; DB 21; Length 74;
 Best Local Similarity 77.8%; Pred. NO. 65;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 II I I I I I
 Db 36 RLRSCKVKV 44

RESULT 30
 ABG17953
 ID ABG17953 standard; Protein; 364 AA.

XX AC ABG17953;
 XX DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #17944.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS82140.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -

XX Claim 20; SEQ ID No 48312; 103pp; English.
 PS The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX Sequence 364 AA;

Query Match 72.5%; Score 29; DB 22; Length 364;
 Best Local Similarity 55.6%; Pred. No. 3.7e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 II: I I I I I
 Db 98 RLAGMVKKRI 106

RESULT 31
 AAW01103
 ID AAW01103 standard; Protein; 582 AA.
 XX AC AAW01103;
 XX DT 20-NOV-1996 (first entry)
 XX DE N. benthamiana phytoene desaturase.
 XX KW phytoene desaturase; Nicotiana; zeta-carotene; carotenoid;
 KW biosynthesis.
 XX OS Nicotiana benthamiana.
 XX PN US5539093-A.
 XX PD 23-JUL-1996.
 XX PF 16-JUN-1994; 94US-0261086.
 XX PR 16-JUN-1994; 94US-0261086.

XX PA (DELL/) DELLA-CIOPPA G R.

XX PA (FITZ/) FITZMAURICE W P.

XX PA (GRILL/) GRILL L K.

XX PA (HELL/) HELLMANN G M.

XX PA (KUMAI/) KUMAGAI M H.

XX PI Della-ciooppa GR, Fitzmaurice WP, Grill LK, Hellmann GM;

XX PI Kumagai MH;

XX DR WPI; 1996-353878/35.

XX DR N-PSDB; AAT40899.

XX PT Phytoene desaturase nucleic acids obtd. from Nicotiana species -

PT used for the biosynthesis or in vitro synthesis of zeta-carotene.

XX Claim 2; Column 13-18; 24pp; English.
XX The present sequence is that of a phytoene desaturase (PD) isolated from
CC Nicotiana benthamiana. The PD is used for the biosynthesis of zeta-
CC carotene which is useful in the carotenoid biosynthesis pathway.
CC Carotenoids such as phytoene have been found to be useful in absorbing
CC ultraviolet radiation, and lycopene has use as a colouring agent in
CC situations in which a red colour is desired.
XX
SQ Sequence 582 AA;

Query Match 72.5%; Score 29; DB 17; Length 582;
Best Local Similarity 55.6%; Pred. No. 6.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||:|:|:
Db 343 RLNSRIKKI 351

RESULT 32
AAW01104
ID AAW01104 standard; Protein; 582 AA.

XX
AC AAW01104;
XX 20-NOV-1996 (first entry)

XX N. benthamiana phytoene desaturase.

DE phytoene desaturase; Nicotiana; zeta-carotene; carotenoid;
XX biosynthesis.

KW Nicotiana benthamiana.

XX US5539093-A.

PN 23-JUL-1996.

XX 16-JUN-1994; 94US-0261086.

PF 16-JUN-1994; 94US-0261086.

XX (DELL/) DELLA-CIOPPA G R.

PA (FITZ/) FITZMAURICE W P.

PA (GRIL/) GRILL L K.

PA (HELL/) HELLMANN G M.

PA (KUMA/) KUMAGAI M H.

XX Della-cioppa GR, Fitzmaurice WP, Grill LK, Hellmann GM;
PI Kumagai MH;

XX WPI; 1996-353878/35.

DR N-PSDB; AAT40700.

XX Phytoene desaturase nucleic acids obtd. from Nicotiana species -
PT used for the biosynthesis or in vitro synthesis of zeta-carotene.

XX Claim 2; Column 19-24; 24pp; English.

XX The present sequence is that of a phytoene desaturase (PD) isolated from
CC Nicotiana benthamiana. The PD is used for the biosynthesis of zeta-
CC carotene which is useful in the carotenoid biosynthesis pathway.
CC Carotenoids such as phytoene have been found to be useful in absorbing
CC ultraviolet radiation, and lycopene has use as a colouring agent in
CC situations in which a red colour is desired.
XX
SQ Sequence 582 AA;

Query Match 72.5%; Score 29; DB 17; Length 582;
Best Local Similarity 55.6%; Pred. No. 6.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||:|:|:
Db 343 RLNSRIKKI 351
RESULT 33
AAW01105
ID AAW01105 standard; Protein; 582 AA.

XX
AC AAW01105;

XX 20-NOV-1996 (first entry)

XX N. tabacum phytoene desaturase.

DE phytoene desaturase; Nicotiana; zeta-carotene; carotenoid;
XX biosynthesis.

XX Nicotiana tabacum.

XX US5539093-A.

PN 23-JUL-1996.

XX 16-JUN-1994; 94US-0261086.

XX 16-JUN-1994; 94US-0261086.

XX (DELL/) DELLA-CIOPPA G R.

PA (FITZ/) FITZMAURICE W P.

PA (GRIL/) GRILL L K.

PA (HELL/) HELLMANN G M.

PA (KUMA/) KUMAGAI M H.

XX Della-cioppa GR, Fitzmaurice WP, Grill LK, Hellmann GM;
PI Kumagai MH;

XX WPI; 1996-353878/35.

DR N-PSDB; AAT40701.

XX Phytoene desaturase nucleic acids obtd. from Nicotiana species -
PT used for the biosynthesis or in vitro synthesis of zeta-carotene.

XX Claim 2; Column 27-32; 24pp; English.

XX The present sequence is that of a phytoene desaturase (PD) isolated from
CC Nicotiana tabacum. The PD is used for the biosynthesis of zeta-carotene
CC which is useful in the carotenoid biosynthesis pathway. Carotenoids such
CC as phytoene have been found to be useful in absorbing ultraviolet
CC radiation, and lycopene has use as a colouring agent in situations in
CC which a red colour is desired.
XX
SQ Sequence 582 AA;

Query Match 72.5%; Score 29; DB 17; Length 582;
Best Local Similarity 55.6%; Pred. No. 6.3e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||:|:|:
Db 343 RLNSRIKKI 351

RESULT 34
AAG33802
ID AAG33802 standard; Protein; 83 AA.

XX
AC AAG33802;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 41024.

XX Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.
 XX EP1033405-A2.
 PN
 XX
 PD 06-SEP-2000.
 XX
 PF 25-FEB-2000; 2000EP-0301439.
 XX
 PR 25-FEB-1999; 99US-0121825.
 PR 05-MAR-1999; 99US-0123180.
 PR 09-MAR-1999; 99US-0123548.
 PR 23-MAR-1999; 99US-0123788.
 PR 29-MAR-1999; 99US-0126264.
 PR 01-APR-1999; 99US-0126785.
 PR 06-APR-1999; 99US-0127462.
 PR 08-APR-1999; 99US-0128234.
 PR 16-APR-1999; 99US-0128714.
 PR 19-APR-1999; 99US-0129845.
 PR 21-APR-1999; 99US-0130077.
 PR 23-APR-1999; 99US-0130449.
 PR 28-APR-1999; 99US-0130510.
 PR 30-APR-1999; 99US-0130891.
 PR 30-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 07-MAY-1999; 99US-0132487.
 PR 11-MAY-1999; 99US-0132863.
 PR 14-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134768.
 PR 19-MAY-1999; 99US-0134941.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 21-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140695.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142390.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 19-JUL-1999; 99US-0144086.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.
 PR 20-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144684.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145085.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145913.
 PR 27-JUL-1999; 99US-0145918.
 PR 28-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 03-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 04-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 05-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 06-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 09-AUG-1999; 99US-0147935.
 PR 10-AUG-1999; 99US-0148171.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 25-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.

PR 31-AUG-1999; 990S-0151438.
 PR 01-SEP-1999; 990S-0151930.
 PR 07-SEP-1999; 990S-0152363.
 PR 10-SEP-1999; 990S-0153070.
 PR 13-SEP-1999; 990S-0153758.
 PR 15-SEP-1999; 990S-0154018.
 PR 16-SEP-1999; 990S-0154039.
 PR 20-SEP-1999; 990S-0154779.
 PR 22-SEP-1999; 990S-0155139.
 PR 23-SEP-1999; 990S-0155486.
 PR 24-SEP-1999; 990S-0155659.
 PR 28-SEP-1999; 990S-0156458.
 PR 29-SEP-1999; 990S-0156596.
 PR 04-OCT-1999; 990S-0157117.
 PR 05-OCT-1999; 990S-0157753.
 PR 06-OCT-1999; 990S-0157865.
 PR 07-OCT-1999; 990S-0158029.
 PR 08-OCT-1999; 990S-0158232.
 PR 12-OCT-1999; 990S-0158369.
 PR 13-OCT-1999; 990S-0159293.
 PR 13-OCT-1999; 990S-0159294.
 PR 13-OCT-1999; 990S-0159295.
 PR 14-OCT-1999; 990S-0159329.
 PR 14-OCT-1999; 990S-0159330.
 PR 14-OCT-1999; 990S-0159331.
 PR 14-OCT-1999; 990S-0159637.
 PR 14-OCT-1999; 990S-0159638.
 PR 18-OCT-1999; 990S-0159584.
 PR 21-OCT-1999; 990S-0160741.
 PR 21-OCT-1999; 990S-0160767.
 PR 21-OCT-1999; 990S-0160768.
 PR 21-OCT-1999; 990S-0160770.
 PR 21-OCT-1999; 990S-0160814.
 PR 21-OCT-1999; 990S-0160815.
 PR 22-OCT-1999; 990S-0160980.
 PR 22-OCT-1999; 990S-0160981.
 PR 22-OCT-1999; 990S-0160989.
 PR 25-OCT-1999; 990S-0161404.
 PR 25-OCT-1999; 990S-0161405.
 PR 25-OCT-1999; 990S-0161406.
 PR 26-OCT-1999; 990S-0161359.
 PR 26-OCT-1999; 990S-0161360.
 PR 26-OCT-1999; 990S-0161361.
 PR 28-OCT-1999; 990S-0161920.
 PR 28-OCT-1999; 990S-0161992.
 PR 28-OCT-1999; 990S-0161993.
 PR 29-OCT-1999; 990S-0162142.

Query Match 70.0%; Score 28; DB 21; Length 83;
 Best Local Similarity 66.7%; Pred. No. 1.2e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMKV 9
 I:II :III
 Db 58 RISSPLKV 66

RESULT 35
 AAO13253
 ID AAO13253 standard; Protein: 87 AA.
 AC AAO13253;
 XX

DT 06-NOV-2001 (first entry)
 XX Human polypeptide SEQ ID NO 27145.
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.

XX WO200164835-A2.
 PN 07-SEP-2001.
 XX 26-FEB-2001; 2001WO-US04927.
 XX 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX (HYSE-) HYSEQ INC.
 XX Tang YT, Liu C, Drmanac RT;
 PI WPI: 2001-514838/56.
 XX N-PSDB; AAI93184.
 DR Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -
 XX Claim 20; SEQ ID NO 27145; 1399pp + Sequence Listing; English.
 CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 87 AA;
 SQ
 Query Match 70.0%; Score 28; DB 22; Length 87;
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMKV 7
 III:III
 Db 74 RLSSMKV 80

RESULT 36
 AAO12825
 ID AAO12825 standard; Protein: 100 AA.
 XX
 AC AAO12825;
 XX

DT 06-NOV-2001 (first entry)
 XX Human polypeptide SEQ ID NO 26717.
 DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 XX vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.
 XX
 OS Homo sapiens.
 XX WO200164835-A2.
 PN 07-SEP-2001.
 XX 26-FEB-2001; 2001WO-US04927.
 XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.
XX (HYSE-) HYSEQ INC.
PA
XX Tang YT, Liu C, Drmanac RT;
PI
XX WPI; 2001-514838/56.
DR N-PSDB; AAI92756.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
XX Claim 20: SEQ ID NO 26717; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO0010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 100 AA;

Query Match 70.0%; Score 28; DB 22; Length 100;
Best Local Similarity 62.5%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKKV 9
I::IIII:
Db 90 LANVKKI 97

RESULT 37
AAW60941
ID AAW60941 standard; Protein; 106 AA.
XX
AC AAW60941;
XX

DT 13-OCT-1998 (first entry)
XX Streptococcus pneumoniae encoded polypeptide.
DE
XX coding region; ORF; open reading frame; antibacterial;
KW infection; prevention; meningitis.
XX

OS Streptococcus pneumoniae.
XX

PN WO9819689-A1.
XX

PD 14-MAY-1998.
XX

PF 27-OCT-1997; 97WO-US19226.
XX

XX 01-NOV-1996; 96US-0029930.
XX

PA (SMIK) SMITHKLINE BEECHAM CORP.
PA (SMIK) SMITHKLINE BEECHAM PLC.

PI Black MT, Hodgson JE, Knowles DJC, Lonetto MA, Nicholas RO;
PI Reid RH, Zarfos PN;
XX

DR WPI: 1998-286586/25.
XX

XX N-PSDB; AAV37341.
XX

PT New isolated nucleic acids from Streptococcus pneumoniae - useful,
PT e.g. for identifying anti-bacterial(s) for treatment and prevention
PT of meningitis
XX

PS Claim 11: Page 44; 130pp; English.
XX

CC The sequence is that of the polypeptide encoded by a region isolated
CC from S. pneumoniae which shows homology to phospho-2-dehydro-3-deoxy-
CC heptonate aldolase. It, or agonists of it, may be useful as an anti-
CC bacterial for treatment or prevention of infection, specifically caused
CC by S.pneumoniae (particularly meningitis) but possibly also Helicobacter
CC pylori (ulcers and gastric cancer). It may be of particular
CC use before insertion of an in-dwelling device or any other
CC invasive procedure. The protein, or nucleic acid encoding
CC it, can also be used in vaccines to induce a cellular
CC and/or humoral immune response, or to screen for other
CC antibacterials. The DNA may also contain flanking sequences
CC that are potential sources of control elements for bacterial
CC gene expression. Detecting a sequence encoding the protein
CC can be used diagnostically, e.g. to detect a mutation for
CC serotyping or classifying infectious agents.
XX

SQ Sequence 106 AA;

Query Match 70.0%; Score 28; DB 19; Length 106;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
I::IIII:
Db 37 RLSALQKKV 45

RESULT 38
AAG33801
ID AAG33801 standard; Protein; 121 AA.
XX
AC AAG33801;
XX

DT 18-OCT-2000 (first entry)
XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 41023.
XX

KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX

OS Arabidopsis thaliana.
XX

PN EP1033405-A2.
XX

PD 06-SEP-2000.
XX

PF 25-FEB-2000; 2000EP-0301439.
XX

XX 25-FEB-1999; 99US-0121825.
PR

PR 05-MAR-1999; 99US-0123180.
PR

PR 09-MAR-1999; 99US-0123548.
PR

PR 23-MAR-1999; 99US-0125788.
PR

PR 25-MAR-1999; 99US-0126264.
PR

PR 29-MAR-1999; 99US-0126785.
PR

PR 01-APR-1999; 99US-0127462.
PR

PR 06-APR-1999; 99US-0128234.
PR

PR 08-APR-1999; 99US-0128714.
PR

PR 16-APR-1999; 99US-0129845.
PR

PR 19-APR-1999; 99US-0130077.
PR

PR 21-APR-1999; 99US-0130449.
PR

PR 23-APR-1999; 99US-0130510.
PR

PR 28-APR-1999; 99US-0130891.
PR

PR 30-APR-1999; 99US-0131449.
PR

PR 30-APR-1999; 99US-0132048.
PR

PR 04-MAY-1999; 99US-0132484.
PR

PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 70.0%; Score 28; DB 21; Length 121;

Best Local Similarity 66.7%; Pred. No. 1.8e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLSSMVKKV 9

Db 96 RISSPLKKV 104

RESULT 39

AAO02532

ID AAO02532 standard; Protein; 168 AA.

XX AC AAO02532;

XX DT 06-NOV-2001 (first entry)

XX DE Human polypeptide SEQ ID NO 16424.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

XX PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX DR WPI: 2001-514838/56.

XX DR N-PSDB; AAI82463.

XX PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders.

XX PS Claim 20; SEQ ID NO 16424; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 168 AA;

Query Match 70.0%; Score 28; DB 22; Length 168;

Best Local Similarity 62.5%; Pred. No. 2.6e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 RLSSMVKK 8

Db 64 RLKSLIKK 71

RESULT 40

AAAR54097

ID AAR54097 standard; Protein; 175 AA.

XX AC AAR54097;

XX DT 01-FEB-1995 (first entry)

XX DE Mouse PAP.

XX KW Pancreatitis-Associated Protein; PAP; mouse; human; bacteria;
 KW aggregation; antibacterial; primer; PCR.

XX OS Mus musculus.

XX PN JP06135998-A.

XX PD 17-MAY-1994.

XX PF 22-OCT-1992; 92JP-0284765.

XX PR 22-OCT-1992; 92JP-0284765.

XX PA (SHIO) SHIONOGI & CO LTD.

XX WPI: 1994-197103/24.

XX DR N-PSDB; AAO64171.

XX PT Human and mouse pancreatitis-associated protein (PAP) - useful
 PT for prepn. of proteinaceous antibacterial agent

XX PS Claim 1-6; Page 5-6; 7pp; Japanese.

XX CC PAP DNA has a high bacteria-aggregating activity and is useful for
 CC the prepn. of PAP which can act as a proteinaceous antibacterial
 CC agent. Total RNA was isolated from mouse. It was hybridised with a
 CC modified salmon spermatozoa DNA and known rat PAP cDNA as a probe.
 CC Two primers are given in AAO64173-74.

SQ Sequence 175 AA;

Query Match 70.0%; Score 28; DB 15; Length 175;

Best Local Similarity 85.7%; Pred. No. 2.7e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 LSSMVKK 8

Db 88 LSSMVKK 94

RESULT 41

AAO06740

ID AAO06740 standard; Protein; 231 AA.

XX AC AAO06740;

XX 17-OCT-2000 (first entry)
DT Arabidopsis thaliana protein fragment SEQ ID NO: 3624.
DE
DE
XX
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145216.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.

```
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 70.0%; Score 28; DB 21; Length 231;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMKVKV 9
Db 147 RLSSMKKM 155

RESULT 42
AAG10371
ID AAG10371 standard; Protein; 231 AA.
AC AAG10371;
XX
DT
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 8665.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
PR 25-FEB-1999; 99US-0121825.
PR 09-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
```

```
PR 28-JUN-1999; 990S-0140823.
PR 29-JUN-1999; 990S-0140991.
PR 30-JUN-1999; 990S-0141287.
PR 01-JUL-1999; 990S-0141842.
PR 01-JUL-1999; 990S-0142154.
PR 02-JUL-1999; 990S-0142055.
PR 06-JUL-1999; 990S-0142390.
PR 08-JUL-1999; 990S-0142803.
PR 09-JUL-1999; 990S-0142920.
PR 12-JUL-1999; 990S-0142977.
PR 13-JUL-1999; 990S-0143542.
PR 14-JUL-1999; 990S-0143624.
PR 15-JUL-1999; 990S-0144005.
PR 16-JUL-1999; 990S-0144085.
PR 16-JUL-1999; 990S-0144086.
PR 19-JUL-1999; 990S-0144325.
PR 19-JUL-1999; 990S-0144331.
PR 19-JUL-1999; 990S-0144332.
PR 19-JUL-1999; 990S-0144333.
PR 19-JUL-1999; 990S-0144334.
PR 19-JUL-1999; 990S-0144335.
PR 20-JUL-1999; 990S-0144352.
PR 20-JUL-1999; 990S-0144632.
PR 20-JUL-1999; 990S-0144884.
PR 21-JUL-1999; 990S-0144814.
PR 21-JUL-1999; 990S-0145086.
PR 21-JUL-1999; 990S-0145088.
PR 22-JUL-1999; 990S-0145085.
PR 22-JUL-1999; 990S-0145087.
PR 22-JUL-1999; 990S-0145089.
PR 22-JUL-1999; 990S-0145192.
PR 23-JUL-1999; 990S-0145145.
PR 23-JUL-1999; 990S-0145218.
PR 23-JUL-1999; 990S-0145224.
PR 26-JUL-1999; 990S-0145276.
PR 27-JUL-1999; 990S-0145913.
PR 27-JUL-1999; 990S-0145918.
PR 27-JUL-1999; 990S-0145919.
PR 28-JUL-1999; 990S-0145951.
PR 02-AUG-1999; 990S-0146386.
PR 02-AUG-1999; 990S-0146388.
PR 02-AUG-1999; 990S-0146389.
PR 03-AUG-1999; 990S-0147038.
PR 04-AUG-1999; 990S-0147204.
PR 04-AUG-1999; 990S-0147302.
PR 05-AUG-1999; 990S-0147192.
PR 05-AUG-1999; 990S-0147260.
PR 06-AUG-1999; 990S-0147303.
PR 06-AUG-1999; 990S-0147416.
PR 09-AUG-1999; 990S-0147493.
PR 09-AUG-1999; 990S-0147935.
PR 10-AUG-1999; 990S-0148171.
PR 11-AUG-1999; 990S-0148319.
PR 12-AUG-1999; 990S-0148341.
PR 13-AUG-1999; 990S-0148565.
PR 13-AUG-1999; 990S-0148684.
PR 16-AUG-1999; 990S-0149368.
PR 17-AUG-1999; 990S-0149175.
PR 18-AUG-1999; 990S-0149426.
PR 20-AUG-1999; 990S-0149722.
PR 20-AUG-1999; 990S-0149723.
PR 20-AUG-1999; 990S-0149929.
PR 23-AUG-1999; 990S-0149902.
PR 23-AUG-1999; 990S-0149930.
PR 25-AUG-1999; 990S-0150566.
PR 26-AUG-1999; 990S-0150884.
PR 27-AUG-1999; 990S-0151065.
PR 27-AUG-1999; 990S-0151066.
PR 27-AUG-1999; 990S-0151080.
PR 30-AUG-1999; 990S-0151303.
PR 31-AUG-1999; 990S-0151438.
PR 01-SEP-1999; 990S-0151930.
PR 07-SEP-1999; 990S-0152363.

PR 10-SEP-1999; 990S-0153070.
PR 13-SEP-1999; 990S-0153758.
PR 15-SEP-1999; 990S-0154018.
PR 16-SEP-1999; 990S-0154039.
PR 20-SEP-1999; 990S-0154779.
PR 22-SEP-1999; 990S-0155139.
PR 23-SEP-1999; 990S-0155486.
PR 24-SEP-1999; 990S-0155659.
PR 28-SEP-1999; 990S-0156458.
PR 29-SEP-1999; 990S-0156596.
PR 04-OCT-1999; 990S-0157117.
PR 05-OCT-1999; 990S-0157753.
PR 06-OCT-1999; 990S-0157865.
PR 07-OCT-1999; 990S-0158029.
PR 08-OCT-1999; 990S-0158232.
PR 12-OCT-1999; 990S-0158369.
PR 13-OCT-1999; 990S-0159293.
PR 13-OCT-1999; 990S-0159294.
PR 13-OCT-1999; 990S-0159295.
PR 14-OCT-1999; 990S-0159329.
PR 14-OCT-1999; 990S-0159330.
PR 14-OCT-1999; 990S-0159331.
PR 14-OCT-1999; 990S-0159637.
PR 14-OCT-1999; 990S-0159638.
PR 18-OCT-1999; 990S-0159584.
PR 21-OCT-1999; 990S-0160741.
PR 21-OCT-1999; 990S-0160767.
PR 21-OCT-1999; 990S-0160768.
PR 21-OCT-1999; 990S-0160770.
PR 21-OCT-1999; 990S-0160814.
PR 22-OCT-1999; 990S-0160815.
PR 22-OCT-1999; 990S-0160980.
PR 22-OCT-1999; 990S-0160981.
PR 22-OCT-1999; 990S-0160989.
PR 25-OCT-1999; 990S-0161404.
PR 25-OCT-1999; 990S-0161405.
PR 25-OCT-1999; 990S-0161406.
PR 26-OCT-1999; 990S-0161359.
PR 26-OCT-1999; 990S-0161360.
PR 26-OCT-1999; 990S-0161361.
PR 28-OCT-1999; 990S-0161920.
PR 28-OCT-1999; 990S-0161992.
PR 28-OCT-1999; 990S-0161993.
PR 29-OCT-1999; 990S-0162142.

Query Match 70.0%; Score 28; DB 21; Length 231;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
Db 147 RLSSMKKM 155
||||:|:|:

RESULT 43
AAG12990
ID AAG12990 standard; Protein; 231 AA.
XX
AC AAG12990;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 12317.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
```

XX 25-FEB-2000; 2000EP-0301439.
PF 99US-0121825.
XX PR 25-FEB-1999;
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130049.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 04-MAY-1999; 99US-0132048.
PR 05-MAY-1999; 99US-0132407.
PR 06-MAY-1999; 99US-0132484.
PR 07-MAY-1999; 99US-0132485.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0132487.
PR 14-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999;
PR 09-JUL-1999; 99US-0142803.
PR 12-JUL-1999; 99US-0142920.
PR 13-JUL-1999; 99US-0142977.
PR 14-JUL-1999; 99US-0143542.
PR 15-JUL-1999; 99US-0143624.
PR 16-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147316.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.

PI 24-SEP-1999; 99US-0155659.
XX 28-SEP-1999; 99US-0156458.
DR 29-SEP-1999; 99US-0156596.
DR 04-OCT-1999; 99US-0157117.
XX 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 22-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 70.0%; Score 28; DB 21; Length 231;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMKVKV 9
 |||||:
Db 147 RLSSMKKM 155

RESULT 44
ABB69690
ID ABB69690 standard; Protein; 261 AA.

AC ABB69690;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 35862.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

OS WO200171042-A2.

PN 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
DR N-PSDB; ABL13793.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
genes from Drosophila and for elucidating cell signalling and cell-cell
interactions -
XX
PS Disclosure; SEQ ID NO 35862; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
capable of detecting 1000 or more genes from Drosophila. The invention is
useful in developmental biology and in elucidating cell signalling and
cell-cell interactions in higher eukaryotes for the development of
insecticides, therapeutics and pharmaceutical drugs. The invention
discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
sequences (ABL01840-ABL16175) and the encoded proteins
(ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 261 AA;

Query Match 70.0%; Score 28; DB 22; Length 261;

Best Local Similarity 62.5%; Pred. No. 4.3e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LSSMKVKV 9

 |||||:
Db 137 LSSMINKI 144

RESULT 45

AAU48010

ID AAU48010 standard; Protein; 328 AA.

XX AC AAU48010;

XX 27-FEB-2002 (first entry)

XX Propionibacterium acnes immunogenic protein #8906.

XX SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;

KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;

KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;

KW dermatological; osteopathic; neuroprotectant.

XX OS Propionibacterium acnes.

XX PN WO200181581-A2.

XX 01-NOV-2001.

XX 20-APR-2001; 2001WO-US12865.

XX 21-APR-2000; 2000US-199047P.

PR 02-JUN-2000; 2000US-208841P.

PR 07-JUL-2000; 2000US-216747P.

XX (CORI-) CORIXA CORP.

XX Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX WPI: 2001-616774/71.

DR N-PSDB; AAS59541.

XX Propionibacterium acnes polypeptides and nucleic acids useful for

PT vaccinating against and diagnosing infections, especially useful for

PT treating acne vulgaris -

XX

PS Example 1; SEQ ID No 9205; 1069pp; English.
 XX
 CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.
 CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 328 AA;

Query Match 70.0%; Score 28; DB 22; Length 328;
 Best Local Similarity 66.7%; Pred. No. 5.5e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSWVKV 9
 ||:|:|:
 DB 190 RLASAVRKV 198

Search completed: November 6, 2002, 12:05:03
 Job time : 26.7778 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 18.4444 Seconds
(without alignments)
84.413 Million cell updates/sec

Title: US-09-905-083-33
Perfect score: 40
Sequence: 1 LLLPLQILL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp Vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query %	Match	Length	ID	Description
1	38	95.0	189	11	Q9D8P8	Q9d8p8 mus musculus
2	38	95.0	198	11	Q9CRE4	Q9cre4 mus musculus
3	38	95.0	201	11	Q9D1M7	Q9d1m7 mus musculus
4	37	92.5	208	8	O21527	O21527 clethrionom
5	36	90.0	146	17	Q9V2D5	Q9v2d5 pyrococcus
6	34	85.0	162	12	Q91T40	Q91t40 lumpy skin
7	34	85.0	162	12	Q91M24	Q91m24 lumpy skin
8	34	85.0	277	17	Q9HP13	Q9hpi3 halobacteri
9	34	85.0	335	6	O97748	O97748 bos taurus
10	34	85.0	845	16	Q97H76	Q97h76 clostridium
11	33	82.5	196	2	Q9X620	Q9x620 salmonella
12	33	82.5	196	11	Q9D226	Q9d226 mus musculus
13	33	82.5	282	16	O53979	O53979 mycobacteri
14	33	82.5	426	16	Q9HW63	Q9hw63 pseudomonas
15	33	82.5	470	8	Q9SEA5	Q9sea5 guillardia
16	33	82.5	475	17	O26479	O26479 methanother

17	32	80.0	150	16	Q9CKE2	Q9cke2 pasteurella
18	32	80.0	198	4	Q13970	Q13970 homo sapien
19	32	80.0	235	4	Q96OR6	Q96or6 homo sapien
20	32	80.0	235	6	Q9XSM6	Q9xsm6 saimiri sci
21	32	80.0	256	11	Q922G5	Q922g5 mus musculus
22	32	80.0	266	11	Q9ESS3	Q9ess3 mus musculus
23	32	80.0	267	11	Q9R1S4	Q9rls4 rattus norv
24	32	80.0	267	11	Q35426	Q35426 mus musculus
25	32	80.0	378	6	Q9N1X4	Q9n1x4 sus scrofa
26	32	80.0	418	4	Q9NW46	Q9nw46 homo sapien
27	32	80.0	612	4	Q9H8Y2	Q9h8y2 homo sapien
28	32	80.0	662	5	Q969D0	Q969d0 giardia lam
29	32	80.0	693	3	Q9C4A2	Q9c4a2 tricholoma
30	32	80.0	1873	10	Q9FG11	Q9fg11 arabidopsis
31	31	77.5	96	12	Q9QU08	Q9qu08 tt virus. o
32	31	77.5	133	6	Q9M2R9	Q9mr29 oryctolagus
33	31	77.5	143	4	O15412	O15412 homo sapien
34	31	77.5	155	4	Q96P81	Q96p81 homo sapien
35	31	77.5	155	11	Q923T2	Q923t2 sigmodon hi
36	31	77.5	159	16	Q9RY14	Q9ry14 deinococcus
37	31	77.5	174	6	Q95LE0	Q95le0 canis famil
38	31	77.5	175	11	Q9D254	Q9d254 mus musculus
39	31	77.5	178	11	Q91X12	Q91x12 cavia porce
40	31	77.5	184	10	Q94IG6	Q94ig6 lithospermu
41	31	77.5	189	5	Q9VYN0	Q9vyn0 drosophila
42	31	77.5	189	8	Q23885	Q23885 dictyosteli
43	31	77.5	204	8	P92567	P92567 arabidopsis
44	31	77.5	227	4	Q9UKJ0	Q9ukj0 homo sapien
45	31	77.5	244	17	O30150	O30150 archaeoglob

ALIGNMENTS

RESULT 1

Q9D8P8	PRELIMINARY;	PRT;	189 AA.
AC Q9D8P8;			
DT 01-JUN-2001 (TrEMBLrel. 17, Created)			
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)			
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE 1110002023RIK PROTEIN.			
GN 1110002023RIK.			
OS Mus musculus (Mouse).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX NCBI_TaxID=10090;			
RN [1]			
RP SEQUENCE FROM N.A.			
RC STRAIN=C57BL/6J; TISSUE=PANCREAS;			
RX MEDLINE=21085660; PubMed=11217851;			
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,			
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,			
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,			
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,			
RA Hayashizaki Y.,			
RT *Functional annotation of a full-length mouse cDNA collection.*;			
RL Nature 409:685-690(2001).			
DR EMBL; AK007813; BAB25278.1; -.			
DR HSSP; P20071; 1FKJ.			

```

DR MGD; MGI:1913370; 1110002023Rik.
DR InterPro: IPR001179; FKBP_PP1ase.
DR Pfam: PF00254; FKBP; 1.
DR PROSITE: PS00453; FKBP_PP1ase_1; UNKNOWN_1.
DR PROSITE: PS00509; FKBP_PP1ase_3; 1.
SQ SEQUENCE 189 AA; 20626 MW; AD979587F1E0582B CRC64;

Query Match
Best Local Similarity 95.0%; Score 38; DB 11; Length 189;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
Db 6 LLLPLQLLL 14

RESULT 2
Q9CRE4
ID Q9CRE4 PRELIMINARY; PRT; 198 AA.
AC Q9CRE4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 1110002023Rik PROTEIN (FRAGMENT).
GN 1110002023Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momaberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK019132; BAB31559.1; -.
DR HSSP; P20071; 1FKJ.
DR MGD; MGI:1913370; 1110002023Rik.
DR InterPro: IPR001179; FKBP_PP1ase.
DR PROSITE: PS00453; FKBP_PP1ase_1; UNKNOWN_1.
DR PROSITE: PS00509; FKBP_PP1ase_3; 1.
DR NON_TER 198
SQ SEQUENCE 198 AA; 21882 MW; A588345383032972 CRC64;

Query Match
Best Local Similarity 95.0%; Score 38; DB 11; Length 198;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
Db 6 LLLPLQLLL 14

RESULT 3
Q9DIM7
ID Q9DIM7 PRELIMINARY; PRT; 201 AA.
AC Q9DIM7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 1110002023Rik PROTEIN.
GN 1110002023Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Momaberts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK019132; BAB31559.1; -.
DR HSSP; P20071; 1FKJ.
DR MGD; MGI:1913370; 1110002023Rik.
DR InterPro: IPR001179; FKBP_PP1ase.
DR PROSITE: PS00453; FKBP_PP1ase_1; UNKNOWN_1.
DR PROSITE: PS00509; FKBP_PP1ase_3; 1.
DR NON_TER 198
SQ SEQUENCE 198 AA; 21882 MW; A588345383032972 CRC64;

Query Match
Best Local Similarity 95.0%; Score 38; DB 11; Length 198;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
Db 6 LLLPLQLLL 14

RESULT 4
Q21527
ID Q21527 PRELIMINARY; PRT; 208 AA.
AC Q21527;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3) (FRAGMENT).
GN ND4.
OS Clethrionomys gapperi (Southern red-backed vole).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;
OC Clethrionomys.
OX NCBI_TaxID=56223;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98152203; PubMed=9491603;
RA Engel S.R., Hogan K.M., Taylor J.F., Davis S.K.;
RT "Molecular systematics and paleobiogeography of the South American
sigmodontine rodents.";

```

RL Mol. Biol. Evol. 15:35-49(1998).
CC -I- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL: U83808; AAB87168.1; -.
DR InterPro: IPR001750; Oxidored_q1.
DR InterPro: IPR000260; Oxidored_q5_N.
DR Pfam: PF00361; oxidored_q1; 1.
DR Pfam: PF01059; oxidored_q5_N; 1.
DR Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
FT NON_TER 208 208
SQ SEQUENCE 208 AA; 23967 MW; 8AF1788697AED6A2 CRC64;

Query Match 92.5%; Score 37; DB 8; Length 208;
Best Local Similarity 88.9%; Pred. No. 7.4;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
||:|||||
Db 98 LLLPLQILL 106

RESULT 5

Q9V2D5 PRELIMINARY; PRT; 146 AA.
AC Q9V2D5;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TRENBLrel. 14, Last annotation update)
DE HYPOTHETICAL 16.1 KDA PROTEIN.
GN PAB0088.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ248283; CAB49063.1; .
KW Hypothetical protein; Complete proteome.

QY SEQUENCE 146 AA; 16092 MW; 7182941371258C1F CRC64;
Query Match 90.0%; Score 36; DB 17; Length 146;
Best Local Similarity 77.8%; Pred. No. 8.4;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||||
Db 67 LLLPLQIII 75

RESULT 6

Q91T40 PRELIMINARY; PRT; 162 AA.
AC Q91T40;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 18.8 KDA PROTEIN.
OS Lumpy skin disease virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Capripoxvirus.
OX NCBI_TaxID=59509;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEETHLING;
RA Stipinovich C., Vreede F.T., Kara P.D., Wallace D.B., Nel L.H.,
RA Viljoen G.J.;
RT "Molecular characterization of important regions of the Lumpy skin disease virus genome.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF336128; AAK43550.1; .

KW Hypothetical protein.
SQ SEQUENCE 162 AA; 18782 MW; 5F914A4080F729EE CRC64;
Query Match 85.0%; Score 34; DB 12; Length 162;
Best Local Similarity 88.9%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||
Db 90 LLLPLTILL 98

RESULT 7

Q91MZ4 PRELIMINARY; PRT; 162 AA.
AC Q91MZ4;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE LSDV010 LAP/PHD-FINGER PROTEIN.
GN LSDV010.
OS lumpy skin disease virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Capripoxvirus.
OX NCBI_TaxID=59509;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEETHLING 2490;
RX MEDLINE=21329495; PubMed=11435593;
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RT "Genome of lumpy skin disease virus.";
RL J. Virol. 75:7122-7130(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=NEETHLING 2490;
RA Tulman E.R., Afonso C.L., Lu Z., Zsak L., Kutish G.F., Rock D.L.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF325528; AAK84971.1; .
SQ SEQUENCE 162 AA; 18835 MW; 0EA24745C3818222 CRC64;

Query Match 85.0%; Score 34; DB 12; Length 162;
Best Local Similarity 88.9%; Pred. No. 24;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||
Db 90 LLLPLTILL 98

RESULT 8

Q9HP13 PRELIMINARY; PRT; 277 AA.
AC Q9HP13;
DT 01-MAR-2001 (TRENBLrel. 16, Created)
DT 01-MAR-2001 (TRENBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TRENBLrel. 16, Last annotation update)
DE CYTOCHROME AA3 CONTROLLING PROTEIN.
GN CCP OR VNG1623G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leitchauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krabs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudis J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.; .

RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL: AE005072; AAG19884.1; -;
KW Complete proteome.
SQ SEQUENCE 277 AA; 29053 MW; B053123765E274B1 CRC64;

Query Match 85.0%; Score 34; DB 17; Length 277;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
:||||:|
DB 97 VLLPQVIL 105

RESULT 9

ID O97748 PRELIMINARY; PRT; 335 AA.
AC O97748;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CONGLUTININ.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93213261; PubMed=8460993;
RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=93277452; PubMed=7684896;
RA Lu J., Laursen S., Thiel S., Jensenius J., Reid K.;
RT "The cDNA cloning of conglutinin and identification of liver as a
primary site of synthesis of conglutinin in members of the Bovidae.";
RL Biochem. J. 292:157-162(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RX MEDLINE=94128104; PubMed=82937370;
RA Kawasaki N., Itoh N., Kawasaki T.;
RT "Gene organization and 5'-flanking region sequence of conglutinin: a
C-type mammalian lectin containing a collagen-like domain.";
RL Biochem. Biophys. Res. Commun. 198:597-604(1994).
DR EMBL: D25302; BAA04983.1; -;
DR EMBL: D25296; BAA04983.1; JOINED.
DR EMBL: D25297; BAA04983.1; JOINED.
DR EMBL: D25299; BAA04983.1; JOINED.
DR EMBL: D25300; BAA04983.1; JOINED.
DR EMBL: D25301; BAA04983.1; JOINED.
DR HSSP: P35247; 1B08.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 2.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE_LLECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE_LLECTIN_2; 1.
FT CONFLICT 236 236 E -> V (IN REF. 2).
SQ SEQUENCE 335 AA; 34702 MW; 39D3A30BC76C134C CRC64;

Query Match 85.0%; Score 34; DB 6; Length 335;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

DB 2 LLLPLSVLL 10
||||:|

RESULT 10

Q97H76
ID Q97H76 PRELIMINARY; PRT; 845 AA.
AC Q97H76;
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CATION TRANSPORT P-TYPE ATPASE.
GN CAC2137.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Oiu D., Hitti J., Wolf Y.I.,
RA Tatusov R.L., Sabath F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007714; AAK80095.1; -;
DR InterPro: IPR004014; Cation_ATPase.
DR InterPro: IPR001757; E1-E2_ATPase.
DR InterPro: IPR000695; HATPase.
DR InterPro: IPR001454; Hydrolase.
DR InterPro: IPR000661; Na_H_K_ATPase.
DR Pfam: PF00689; Cation_ATPase_C; 1.
DR Pfam: PF00690; Cation_ATPase_N; 1.
DR Pfam: PF00122; E1-E2_ATPase; 1.
DR Pfam: PF00702; Hydrolase; 1.
DR PRINTS: PR00119; CATATPASE.
DR PRINTS: PR00120; HATPASE.
DR PROSITE: PS00154; ATPASE_E1_E2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 845 AA; 93779 MW; 401293AA1FF9D757 CRC64;

Query Match 85.0%; Score 34; DB 16; Length 845;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
||||:|
DB 670 LLLPLQILL 677

RESULT 11

ID Q9X620 PRELIMINARY; PRT; 196 AA.
AC Q9X620;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE CORF.
GN CORF.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RA Smith R.L., Ahuga D., Thacker L.K., Maguire M.E.;
RT "Magnesium transport in Salmonella typhimurium: Sequence and
characterization of the corB, corC, and corD genes.";
RL Submitted (FEB-1999) to the EMBL/GenBank/DBSJ databases.

```

DR EMBL; AF130857; AAD31438.1; -.
DR InterPro; IPR002550; DUF21.
DR Pfam; PF01595; DUF21.1.
SQ SEQUENCE 196 AA; 21859 MW; B06659F7E5BAA17A CRC64;

Query Match      82.5%; Score 33; DB 2; Length 196;
Best Local Similarity 77.8%; Pred. No. 46;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 123 LLAPLQILM 131

RESULT 12
Q9D226 PRELIMINARY; PRT; 196 AA.
AC Q9D226:
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE A030007E19RIK PROTEIN.
GN A030007E19RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SKIN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohetsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK020699; BAB32180.1; -.
DR HSSP; P10969; 1WGT.
DR MGD; MGI:1924465; A030007E19RIK.
DR InterPro; IPR001450; 4FE4S_ferredoxin.
DR PROSITE; PS00198; 4FE4S_FERREDOXIN; UNKNOWN.1.
SQ SEQUENCE 196 AA; 17777 MW; 2136738FD496D815 CRC64;

Query Match      82.5%; Score 33; DB 11; Length 196;
Best Local Similarity 77.8%; Pred. No. 46;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 156 LLLPAQLLL 164

RESULT 13
Q53979 PRELIMINARY; PRT; 282 AA.
AC Q53979:
DT 01-JUN-1998 (TREMBLrel. 06, Created)
DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)

```

```

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 30.6 KDA PROTEIN.
GN RV1978 OR MTV051.16.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; AL022073; CAAL7851.1; -.
DR TubercuList; RV1978; -.
DR InterPro; IPR000051; SAM_bind.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 282 AA; 30584 MW; 3CA18BD208951D1E CRC64;

Query Match      82.5%; Score 33; DB 16; Length 282;
Best Local Similarity 77.8%; Pred. No. 65;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 196 LLLPLHLL 204

RESULT 14
Q5HW63 PRELIMINARY; PRT; 426 AA.
AC Q5HW63:
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE HYPOTHETICAL PROTEIN PA4338.
GN PA4338.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Miziochichi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004850; AAG07726.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 426 AA; 48323 MW; AB00AA9643AF5257 CRC64;

Query Match      82.5%; Score 33; DB 16; Length 426;
Best Local Similarity 77.8%; Pred. No. 97;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 1111 I:111

```

Db 301 LLLPAQLLL 309

RESULT 15

Q9SEA5 PRELIMINARY; PRT; 470 AA.

AC Q9SEA5;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE HYPOTHETICAL 57.8 KDA PROTEIN.

GN ORF470.

OS Guillardia theta (Cryptomonas phi).

OG Nucleomorph.

OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.

OX NCBI_TaxID=55529;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20087226; PubMed=10618395;

RA Zauner S., Fraunholz M., Wastl J., Penny S., Beaton M., Cavalier-Smith T., Maier U.G., Douglas S.;

RT "Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph.";

RL Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=21223349; PubMed=11323671;

RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T., Wu X., Reith M., Cavalier-Smith T., Maier U.G.;

RT "The highly reduced genome of an enslaved algal nucleus.";

RL Nature 410:1091-1096(2001).

DR EMBL: AF165818; AAF24211.1; -

KW Hypothetical protein.

SQ SEQUENCE 470 AA; 57801 MW; 19FAA35E26A452FA CRC64;

Query Match 82.5%; Score 33; DB 8; Length 470;

Best Local Similarity 66.7%; Pred. No. 1.1e+02;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

Db 253 MLLPLEILI 261

RESULT 16

O26479 PRELIMINARY; PRT; 475 AA.

AC O26479;

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)

DE O-ANTIGEN TRANSPORTER RELATED PROTEIN.

GN MTH379.

OS Methanothermobacter thermoautotrophicus.

OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;

OX Methanothermobacter.

OX NCBI_TaxID=145262;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DELTA H;

RX MEDLINE=98037514; PubMed=9371463;

RA Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J., Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K., Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D., Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R., Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S., McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M., Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;

RT "Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics.";

RL J. Bacteriol. 179:7135-7155(1997).

DG EMBL: AE000823; AAB84885.1; -

DR InterPro; IPR002797; Polysacc_synt.

DR InterPro; IPR001680; WD40.

DR Pfam; PF01943; Polysacc_synt; 1.

DR PROSITE; PS00678; WD_REPEATS_1; UNKNOWN_1.

KW Complete proteome.

SQ SEQUENCE 475 AA; 51904 MW; 97852B34928608DD CRC64;

Query Match 82.5%; Score 33; DB 17; Length 475;

Best Local Similarity 77.8%; Pred. No. 1.1e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

Db 432 LLLPLSILI 440

RESULT 17

Q9CKE2 PRELIMINARY; PRT; 150 AA.

ID Q9CKE2

AC Q9CKE2

DT 01-JUN-2001 (TREMBlrel. 17, Created)

DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE HYPOTHETICAL PROTEIN PM1680.

GN PM1680.

OS Pasteurella multocida.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Pasteurella.

OX NCBI_TaxID=747;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PM70;

RX MEDLINE=21145866; PubMed=11248100;

RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;

RT "Complete genomic sequence of Pasteurella multocida Pm70.";

RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).

DR EMBL: AE006204; AAK03764.1; -

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 150 AA; 17314 MW; 7454710498463FF9 CRC64;

Query Match 80.0%; Score 32; DB 16; Length 150;

Best Local Similarity 66.7%; Pred. No. 57;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

Db 24 LLLPLQIVI 32

RESULT 18

Q13970 PRELIMINARY; PRT; 198 AA.

ID Q13970

AC Q13970;

DT 01-NOV-1996 (TREMBlrel. 01, Created)

DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE CD8 ANTIGEN.

GN CD8A.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89215302; PubMed=2496167;

RA Norment A.M., Lonberg N., Lacy E., Littman D.R.;

RT "Alternatively spliced mRNA encodes a secreted form of human CD8 alpha. Characterization of the human CD8 alpha gene.";

RL J. Immunol. 142:3312-3319(1989).

DR EMBL: M26315; AAA79218.1; -

DR EMBL: M26313; AAA79218.1; JOINED.

DR EMBL: M26314; AAA79218.1; JOINED.

DR HSSP; P01732; 1CD8.

DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; ig; 1.
DR SMART: SM00406; IGV; 1.
SQ SEQUENCE 198 AA; 21585 MW; BC89FA59F150598F CRC64;

Query Match 80.0%; Score 32; DB 4; Length 198;

Best Local Similarity 77.8%; Pred. No. 74;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||:|

Db 8 LLLPLALLL 16

RESULT 19

Q96QR6 PRELIMINARY; PRT; 235 AA.

AC Q96QR6;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE MUTANT CD8 ALPHA ANTIGEN.

GN CD8A.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=21329138; PubMed=11435463;

RA de La Calle-Martin O., Hernandez M., Ordi J., Casamitjana N.,

RA Arostegui J.I., Caragol I., Ferrando M., Labrador M.,

RA Rodriguez-Sanchez J.L., Espanol F.,

RT "Familial CD8 deficiency due to a mutation in the CD8alpha gene."

RL J. Clin. Invest. 108:117-117(2001).

DR EMBL; AY039664; AAK72403.1; -.

SQ SEQUENCE 235 AA; 25759 MW; FCCA347AAEF732BB CRC64;

Query Match 80.0%; Score 32; DB 4; Length 235;

Best Local Similarity 77.8%; Pred. No. 88;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||:|

Db 8 LLLPLALLL 16

RESULT 20

Q9XSM6 PRELIMINARY; PRT; 235 AA.

AC Q9XSM6;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE CD8 ALPHA CHAIN PRECURSOR.

OS Saimiri sciureus (Common squirrel monkey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.

OX NCBI_TaxID=9521;

RN [1]

RP SEQUENCE FROM N.A.

RX STRAIN=92039;

RC MEDLINE=99299362; PubMed=10369934;

RA Ureta-Vidal A., Garcia Z., Lemonnier F.a., Kazanji M.;

RT "Molecular characterization of cDNAs encoding squirrel monkey (Saimiri

RL sciureus) CD8 alpha and beta chains.";

RL Immunogenetics 49:718-721(1999).

DR EMBL; A0130818; CAB41462.1; -.

DR HSSP; P01732; ICD8.

DR InterPro: IPR003599; Ig.

DR InterPro: IPR003006; Ig_MHC.

DR InterPro: IPR002965; P_rich_extensn.

DR Pfam: PF00047; ig; 1.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART: SM00409; IG; 1.
KW Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 235 CD8 ALPHA CHAIN.
SQ SEQUENCE 235 AA; 25728 MW; 055867CD503C268D CRC64;

Query Match 80.0%; Score 32; DB 6; Length 235;

Best Local Similarity 77.8%; Pred. No. 88;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||:|

Db 8 LLLPLALLL 16

RESULT 21

Q922G5 PRELIMINARY; PRT; 256 AA.

AC Q922G5;

DT 01-DEC-2001 (TrEMBLrel. 19, Created)

DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE SIMILAR TO X-BOX BINDING PROTEIN 1 (FRAGMENT).

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Strausberg R.;

RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC008153; AA08153.1; -.

FT NON_TER 1

SQ SEQUENCE 256 AA; 28651 MW; 5544DB56D7B2620 CRC64;

Query Match 80.0%; Score 32; DB 11; Length 256;

Best Local Similarity 100.0%; Pred. No. 95;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPLQIL 8

|||||

Db 177 LLPLQIL 183

RESULT 22

Q9ESS3 PRELIMINARY; PRT; 266 AA.

AC Q9ESS3;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE TAX-RESPONSIVE ELEMENT-BINDING PROTEIN 5.

GN TREB5.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20363095; PubMed=10907849;

RA Masaki T., Noguchi H., Kobayashi M., Yoshida M., Takamatsu K.;

RT "Isolation and characterization of the gene encoding mouse tax-

RT responsive element-binding protein (TREB)5."

RL DNA Res. 7:187-193(2000).

CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.

DR EMBL; AB036745; BAB13793.1; -.

DR InterPro: IPR001871; bzip.

DR SMART; SM00338; BRLZ; 1.

DR PROSITE; PS00036; BZIP_BASIC; 1.

DR DNA-Binding; Nuclear protein.

SQ SEQUENCE 266 AA; 29562 MW; 4161FFA93B1800A5 CRC64;

Query Match 80.0%; Score 32; DB 11; Length 266;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPLQL 8

Db 187 LLPLQL 193

RESULT 23

Q9R1S4
ID Q9R1S4 PRELIMINARY; PRT; 267 AA.

AC Q9R1S4;

DT 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE HEPATOCARCINOGENESIS-RELATED TRANSCRIPTION FACTOR (HTF).

GN HTF.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WISTAR;

RX MEDLINE=20137507; PubMed=10675042;

RA Kokura K., Kishimoto T., Tamura T.;

RT "Identity between rat htF and human xbp-1 genes: determination of gene structure, target sequence, and transcription promotion function for HTF.";

RL Gene 241:297-307(2000).

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.

DR EMBL; AB030238; BAA82600.1; -.

DR InterPro; IPR001871; bZIP.

DR Pfam; PF00170; bZIP; 1.

DR SMART; SM00338; BRLZ; 1.

DR PROSITE; PS00036; BZIP_BASIC; 1.

KW DNA-binding, Nuclear protein.

SQ SEQUENCE 267 AA; 29665 MW; B5A58F1D3FAA10B4 CRC64;

Query Match

Best Local Similarity 100.0%; Score 32; DB 11; Length 267;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPLQL 8

Db 188 LLPLQL 194

RESULT 24

O35426
ID O35426 PRELIMINARY; PRT; 267 AA.

AC O35426;

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE X BOX BINDING PROTEIN-1.

GN XBPI.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Lee C.M., Reddy E.P.;

RT "Sequence Analysis of Murine XBP-1.";

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.

DR EMBL; AF027963; AAB81862.2; -.

DR MGD; MGI:98970; Xbp1.
DR InterPro; IPR001871; bZIP.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW DNA-binding, Nuclear protein.
SQ SEQUENCE 267 AA; 29619 MW; 15009E684F6D426F CRC64;

Query Match 80.0%; Score 32; DB 11; Length 267;
Best Local Similarity 100.0%; Pred. No. 99;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPLQL 8

Db 188 LLPLQL 194

RESULT 25

Q9N1X4

ID Q9N1X4 PRELIMINARY; PRT; 378 AA.

AC Q9N1X4;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)

DE LUNG SURFACTANT PROTEIN D PRECURSOR.

GN SFTPD.

OS Sus scrofa (Pig).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

OX NCBI_TaxID=9823;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20109098; PubMed=10640760;

RA van Eijk M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,

RA Lawson P.R.;

RT "Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal

RT localisation and tissue distribution.";

RL J. Immunol. 164:1442-1450(2000).

DR EMBL; AF132496; AAF22145.2; -.

DR HSSP; P35247; 1B08.

DR InterPro; IPR000087; Collagen.

DR InterPro; IPR001304; lectin_c.

DR Pfam; PF01391; Collagen; 3.

DR Pfam; PF00059; lectin_c; 1.

DR SMART; SM00034; CLECT; 1.

DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.

DR PROSITE; PS50041; C_TYPE_LLECTIN_2; 1.

KW Signal.

FT SIGNAL 1 20 POTENTIAL.

FT CHAIN 21 378 LUNG SURFACTANT PROTEIN D.

SQ SEQUENCE 378 AA; 37986 MW; 3504E8C1E56C341D CRC64;

Query Match 80.0%; Score 32; DB 6; Length 378;

Best Local Similarity 66.7%; Pred. No. 1.4e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQL 9

Db 2 LLPLSVLI 10

RESULT 26

Q9NW46

ID Q9NW46 PRELIMINARY; PRT; 418 AA.

AC Q9NW46;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)

DE CDNA FLJ10320 FIS, CLONE NT2RM2000502.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Wagatsuma M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakanatsu A., Nakamura Y., Nagahari K., Masuho Y.,
RA Ninomiya K., Iwayanagi T.,
RT "NEO human cDNA sequencing project."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK001182; BAA91540.1; -.
SQ SEQUENCE 418 AA; 44861 MW; B85F301A96DD545F CRC64;

Query Match 80.0%; Score 32; DB 4; Length 418;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9
||||| :||
Db 398 LLLPLPLL 406

RESULT 27
Q9H8Y2 PRELIMINARY; PRT; 612 AA.
AC Q9H8Y2;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CDNA FLJ13150 FIS, CLONE NT2RP3003353, WEAKLY SIMILAR TO HYPOTHETICAL
DE 26.2 KDA PROTEIN IN GDII1-COX15 INTERGENIC REGION.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
RA Tanase T., Nomura Y., Togiya S., Komai F., Hara K., Takeuchi K.,
RA Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
RA Wakamatsu A., Nakamura Y., Nagahari K., Masuho Y., Oshima A.,
RT "NEO human cDNA sequencing project."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK023212; BAB14465.1; -.
SQ SEQUENCE 612 AA; 69448 MW; E267A5B6EADB5C86 CRC64;

Query Match 80.0%; Score 32; DB 4; Length 612;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9
||||| :||
Db 517 LLVPLQITL 525

RESULT 28
Q969D0 PRELIMINARY; PRT; 662 AA.
AC Q969D0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KINESIN-LIKE PROTEIN 1 (FRAGMENT).
GN GLKPL.
OS Giardia lamblia (Giardia intestinalis).
OC Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX NCBI_TaxID=5741;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=POLISH;
RA Iwabe N.;
RT "Giardia lamblia kinesin-like protein.";

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB028051; BAB56138.1; -.
FT NON_TER 1
SQ SEQUENCE 662 AA; 72349 MW; 5D215D8F9634213A CRC64;

Query Match 80.0%; Score 32; DB 5; Length 662;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQI 7
|||||
Db 82 LLLPLQI 88

RESULT 29
Q9C4A2 PRELIMINARY; PRT; 693 AA.
ID Q9C4A2;
AC Q9C4A2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE POL-LIKE PROTEIN POL-2.
OS Tricholoma matsutake.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Tricholomataceae; Tricholoma.
OX NCBI_TaxID=40145;
RN [1]
RP SEQUENCE FROM N.A.
RA Murata H., Miyazaki Y., Yamada A.;
RT "marY2N, a LINE-like non-long terminal repeat (non-LTR) retroelement
RT from the ectomycorrhizal homobasidiomycete Tricholoma matsutake.";
RL Biosci. Biotechnol. Biochem. 0:0-0(2001).
DR EMBL: AB047280; BAB32470.1; -.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00078; rvt; 1.
DR RNA-directed DNA polymerase.
SQ SEQUENCE 693 AA; 76528 MW; 137015E7E7EB7BE9 CRC64;

Query Match 80.0%; Score 32; DB 3; Length 693;
Best Local Similarity 75.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQLL 9
|||:|:|
Db 493 LLPWQLLL 500

RESULT 30
Q9FGI1 PRELIMINARY; PRT; 1873 AA.
ID Q9FGI1;
AC Q9FGI1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 5, p1 CLONE:MOJ2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsiis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB025632; BAB10256.1; -.
DR EMBL: AB019228; BAB10256.1; JOINED.
DR InterPro; IPR001841; znf_ring.
DR SMART; SM00184; RING; 1.
SQ SEQUENCE 1873 AA; 208743 MW; 777F4E6E620854EA CRC64;

Query Match 80.0%; Score 32; DB 10; Length 1873;
Best Local Similarity 75.0%; Pred. No. 6.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPLOIILL 9
|||||:|
Db 1194 LLPLEVLL 1201

RESULT 31

O9QU08 PRELIMINARY; PRT; 96 AA.
AC Q9QU08;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ORF2 PROTEIN (FRAGMENT).
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TTV BK1024;
RX MEDLINE=20129505; PubMed=10664406;
RA Mulyanto, Hijikata M., Matsushita M., Ingkokusmo G., Widjaya A.,
RA Sumarsidi D., Kanai K., Ohta Y., Mishiro S.;
RT "TT virus (TTV) genotypes in native and non-native prostitutes of
RT Irian Jaya, Indonesia: implication for non-occupational
RT transmission.";
RL Arch. Virol. 145:63-72(2000).
DR EMBL; AB031685; BAA86148.1; -.
DR InterPro; IPR004118; TT_0RF2.
DR Pfam; PF02957; TT_0RF2; 1.
FT NON_TER 1
ET NON_TER 96
SQ SEQUENCE 96 AA; 11069 MW; FC006B668AC1F8F1 CRC64;

Query Match 77.5%; Score 31; DB 12; Length 96;
Best Local Similarity 87.5%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLPLOIILL 9
||||| |
Db 13 LLPLOIILL 20

RESULT 32

Q9MZ9 PRELIMINARY; PRT; 133 AA.
AC Q9MZ9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE INTERLEUKIN 2 VARIANT IL2DELTA2.
GN IL-2.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SPLEEN, LYMPH NODE;
RX MEDLINE=20304414; PubMed=10843729;
RA Perkins H.D., van Leeuwen B.H., Hardy C.M., Kerr P.J.;
RT "The complete cDNA sequences of IL-2, IL-4, IL-6 and IL-10 from the
RT European rabbit (Oryctolagus cuniculus).";
RL Cytokine 12:555-565(2000).
DR EMBL; AF169168; AAF86652.1; -.
DR HSSP; P01585; 3INK.
DR InterPro; IPR000779; Interleukin-2.
DR Pfam; PF00715; IL2; 1.
DR ProDom; PD003649; Interleukin-2; 1.
DR SMART; SM00189; IL2; 1.

DR PROSITE; PS00424; INTERLEUKIN_2; 1.
SQ SEQUENCE 133 AA; 14748 MW; OD54758C190B5655 CRC64;

Query Match 77.5%; Score 31; DB 6; Length 133;
Best Local Similarity 77.8%; Pred. No. 81;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLOIILL 9
|||||:|
Db 37 LLLDLQVLL 45

RESULT 33

O15412 PRELIMINARY; PRT; 143 AA.
AC O15412;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CTG4A.
GN CTG4A.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN CORTEX;
RX MEDLINE=97369492; PubMed=9225980;
RA Margolis R.L., Abraham M.R., Gatchell S.B., Li S.H., Kidwai A.S.,
RA Breschel T.S., Stine O.C., Callahan C., McInnis M.G., Ross C.A.;
RT "cDNAs with long CAG trinucleotide repeats from human brain.";
RL Hum. Genet. 100:114-122(1997).
DR EMBL; U80744; AAB91442.1; -.
SQ SEQUENCE 143 AA; 15959 MW; OF2119BF4B33A1E6 CRC64;

Query Match 77.5%; Score 31; DB 4; Length 143;
Best Local Similarity 77.8%; Pred. No. 87;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLOIILL 9
|||||:|
Db 13 LLLPLLLL 21

RESULT 34

Q96P81 PRELIMINARY; PRT; 155 AA.
AC Q96P81;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PTPRE (FRAGMENT).
GN PTPRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RA Zhao Y., Sun D., Dai S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF406557; AAL01375.1; -.
FT NON_TER 155
SQ SEQUENCE 155 AA; 17359 MW; 8A7DB29677F06E00 CRC64;

Query Match 77.5%; Score 31; DB 4; Length 155;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLOIILL 9
|||||:|

```
Db 51 LLLPLLLLL 59

RESULT 35
Q923T2
ID Q923T2 PRELIMINARY; PRT; 155 AA.
AC Q923T2;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE INTERLEUKIN 2.
OS Sigmodon hispidus (hispid cotton rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Sigmodontinae;
OC Sigmodon.
OX NCBI_TaxID=42415;
RN [1]
RP SEQUENCE FROM N.A.
RA Darnell M.R., Pletneva L.M., Langley R.J., Blanco J.C., Prince G.A.;
RT "Cloning, expression and purification of cotton rat IL-2.";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF398549; AAK94012.1; -.
SQ SEQUENCE 155 AA; 17627 MW; ACADEA865E993291 CRC64;

Query Match 77.5%; Score 31; DB 11; Length 155;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
   ||| |||
Db 37 LLLDLQVLL 45

RESULT 36
Q9RY14
ID Q9RY14 PRELIMINARY; PRT; 159 AA.
AC Q9RY14;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TREMBLrel. 16, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN DR0138.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN SEQUENCE FROM N.A.
RP STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567366;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Morfat K.S., Qin H., Jiang L., Panphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL: AE001876; AAF09727.1; -.
DR TIGR: DR0138; -.
KW Complete proteome.
SQ SEQUENCE 159 AA; 16699 MW; 86D1E17E3E5A02C8 CRC64;

Query Match 77.5%; Score 31; DB 16; Length 159;
Best Local Similarity 66.7%; Pred. No. 96;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
   :||| |||
Db 120 ILLPLSLLL 128
```

```
RESULT 37
Q95LE0
ID Q95LE0 PRELIMINARY; PRT; 174 AA.
AC Q95LE0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE E-CADHERIN (FRAGMENT).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21324723; PubMed=11430829;
RA Comijn J., Bex G., Vermassen P., Verschuere K., van Grunsven L.,
RA Bruyneel E., Mareel M., Huylebrouck D., van Roy F.;
RT "The two-handed E box binding zinc finger protein S1P1 downregulates
E-cadherin and induces invasion.";
RL Mol. Cell 7:1267-1278(2001).
DR EMBL: AF330162; AAL09464.1; -.
FT CHAIN 157 >174 E-CADHERIN.
FT NON_TER 174 174
SQ SEQUENCE 174 AA; 19520 MW; EA40F4775B59B0FD CRC64;

Query Match 77.5%; Score 31; DB 6; Length 174;
Best Local Similarity 77.8%; Pred. No. 11e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
   ||||| |||
Db 11 LLLPLLLLL 19

RESULT 38
Q9D254
ID Q9D254 PRELIMINARY; PRT; 175 AA.
AC Q9D254;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ADULT MALE DIENCEPHALON CDNA, RIKEN FULL-LENGTH ENRICHED LIBRARY,
DE CLONE:9330159H09, FULL INSERT SEQUENCE.
GN TNFRSF11A.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=DIENCEPHALON;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hall D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
EMBL: AK020374; BAB32085.1; -.
```

```

DR HSP; P25942; LCDF.
DR MGD; MG1:1314891; Tnfrs11a.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR001368; TNFR_C6.
DR Pfam; PF00020; TNFR_C6; 3.
DR ProDom; PD000771; TNFR_C6; 1.
DR SMART; SM00208; TNFR; 3.
DR PROSITE; PS00290; IG_MHC; UNKNOWN_1.
DR PROSITE; PS00500; TNFR_NGFR_2; 1.
SQ SEQUENCE 175 AA; 19266 MW; 5DF4846253B11719 CRC64;

Query Match          77.5%; Score 31; DB 11; Length 175;
Best Local Similarity 66.7%; Pred. NO. 1.le+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 21 LLVPLQVTL 29

RESULT 39
Q91X12 PRELIMINARY; PRT: 178 AA.
AC Q91X12;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NEUTROPHIL CATIONIC ANTIBACTERIAL POLYPEPTIDE OF 11 KDA.
GN CAP11.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HARTLEY; TISSUE=BONE MARROW;
RX MEDLINE=97426420; PubMed=9278433;
RA Nagaoaka I., Tsutsumi-Ishii Y., Yomogida S., Yamashita T.;
RT "Isolation of cDNA encoding guinea pig neutrophil cationic
RT antibacterial polypeptide of 11 kDa (CAP11) and evaluation of CAP11
RT mRNA expression during neutrophil maturation.";
RL J. Biol. Chem. 272:22742-22750(1997).
DR EMBL; D87405; BAB60811.1; -.
SQ SEQUENCE 178 AA; 20132 MW; FAC32418246F611F CRC64;

Query Match          77.5%; Score 31; DB 11; Length 178;
Best Local Similarity 77.8%; Pred. NO. 1.le+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 14 LLLPLLLLL 22

RESULT 40
Q94IG6 PRELIMINARY; PRT: 184 AA.
AC Q94IG6;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 21.2 KDA PROTEIN.
GN LEPS-2.
OS Lithospermum erythrorhizon.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; euasterids I incertae sedis; Boraginaceae;
OC Lithospermum.
OX NCBI_TaxID=34254;
RN [1]
RP SEQUENCE FROM N.A.
RA Yamamura Y., Sahin F.P., Mizukami H.;
RT "Identification and characterization of a gene differentially

```

```

RT expressed in a shikonin-producing cell line of Lithospermum
RT erythrorhizon in culture by a fluorescent differential display
RT technique.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB047916; BAB61106.1; -.
KW Hypothetical protein.
SQ SEQUENCE 184 AA; 21207 MW; 00AD51D2110621EE CRC64;

Query Match          77.5%; Score 31; DB 10; Length 184;
Best Local Similarity 87.5%; Pred. NO. 1.le+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
Db 8 LLLPLSILL 15

RESULT 41
Q9VYNO PRELIMINARY; PRT: 189 AA.
AC Q9VYNO;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE CG15927 PROTEIN.
GN CG15927.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikkos G.L.G.,
RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacle J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu G., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).

```

DR EMBL: AE003489; AAF48162.1; -.
DR FlyBase: FBgn0030389; CGI5927.
SQ SEQUENCE 189 AA; 22107 MW; 823E42FAF613F4BF CRC64;

Query Match 77.5%; Score 31; DB 5; Length 189;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQIL 8

|||||

Db 42 LLLPLNIL 49

RESULT 42

O23885
ID Q23885 PRELIMINARY; PRT; 189 AA.
AC Q23885; Q9XPJ5.
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE ORF189.
OS Dictyostelium discoideum (Slime mold).
OG Mitochondrion.
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=AX3, AND PARTIALLY X22 (48172-51517 BP);
RX MEDLINE=96038265; PubMed=8581739;
RA Ogawa S., Yoshino R., Angata K., Pi M., Iwamoto M., Kuroe K.,
RA Matsuo K., Morio T., Urushihara H., Yanagisawa K., Tanaka Y.;
RT "Mitochondrial ribosomal protein L11 gene of Dictyostelium discoideum
resides not in the nuclear genome but in the mitochondrial genome.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.

RC STRAIN=AX3, AND PARTIALLY X22 (48172-51517 BP);
RX MEDLINE=96038265; PubMed=8581739;
RA Ogawa S., Yoshino R., Angata K., Pi M., Iwamoto M., Kuroe K.,
RA Matsuo K., Morio T., Urushihara H., Yanagisawa K., Tanaka Y.;
RT "Mitochondrial ribosomal protein L11 gene of Dictyostelium discoideum
resides not in the nuclear genome but in the mitochondrial genome.";
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3, AND PARTIALLY X22 (48172-51517 BP);
RX MEDLINE=98228272; PubMed=9560439;
RA Iwamoto M., Pi M., Kurihara M., Morio T., Tanaka Y.;
RT "A ribosomal protein gene cluster is encoded in the mitochondrial DNA
of Dictyostelium discoideum: UGA termination codons and similarity of
the gene order to Acanthamoeba castellanii.";
RL Curr. Genet. 33:304-310(1998).
DR EMBL: D21196; BAA04734.1; -.
DR EMBL: AB000109; BAA78070.1; -.
KW Mitochondrion.
SQ SEQUENCE 189 AA; 21866 MW; 50B69EA533A471DB CRC64;

Query Match 77.5%; Score 31; DB 8; Length 189;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQIL 9

|||||

Db 106 ILPLQILV 113

RESULT 43

P92567
ID P92567 PRELIMINARY; PRT; 204 AA.
AC P92567;
DT 01-MAY-1997 (TREMBLrel. 03, Created)
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ORF204.

OS Arabidopsis thaliana (Mouse-ear cress).
OG Mitochondrion.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsiis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Unseld M., Marienfeld J.R., Brandt P., Brennicke A.;
RL Nat. Genet. 0:0-0(0).
DR EMBL: Y08502; CAA69788.1; -.
KW Mitochondrion.
SQ SEQUENCE 204 AA; 22814 MW; E8539AFF8C36458C CRC64;

Query Match 77.5%; Score 31; DB 8; Length 204;
Best Local Similarity 62.5%; Pred. No. 1.2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQIL 9

|||||

Db 163 LIPIQVLL 170

RESULT 44

Q9UKJ0
ID Q9UKJ0 PRELIMINARY; PRT; 227 AA.
AC Q9UKJ0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE ACTIVATING RECEPTOR PILRBETA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20127940; PubMed=10660620;
RA Mousseau D.D., Banville D., L'Abbe D., Bouchard P., Shen S.H.;
RT "PILRBeta, a novel immunoreceptor tyrosine-based inhibitory motif-
bearing protein, recruits SHP-1 upon tyrosine phosphorylation and is
paired with the truncated counterpart PILRBeta.";
RL J. Biol. Chem. 275:4467-4474(2000).
DR EMBL: AF161081; AAD52965.1; -.
DR InterPro: IPR003599; Ig.
DR SMART: SM00409; IG; 1.
KW Receptor.
SQ SEQUENCE 227 AA; 25542 MW; 7FF960C60AB7EF09 CRC64;

Query Match 77.5%; Score 31; DB 4; Length 227;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||

Db 5 LLLPLLLLL 13

RESULT 45

O30150
ID O30150 PRELIMINARY; PRT; 244 AA.
AC O30150;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NITRATE ABC TRANSPORTER, PERMEASE PROTEIN (NRTB-1).
GN AF0086.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
OX NCBI_TaxID=2234;
RN [1]
RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE=98049343; PubMed=9389475;
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,
RA Richardson D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artach P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon *Archaeoglobus fulgidus*.";
RL Nature 390:364-370(1997).
DR EMBL: AE001100; AAB91143.1; -.
DR TIGR: AF0086; -.
DR InterPro: IPR000515; BPD_transp.
DR Pfam: PF00528; BPD_transp; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 244 AA; 26825 MW; BB789B5EBBF8538 CRC64;

Query Match 77.5%; Score 31; DB 17; Length 244;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9
Db 102 VLLPLYL 110

Search completed: November 6, 2002, 12:12:01
Job time : 21.4444 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 5.33333 Seconds
(without alignments)
65.339 Million cell updates/sec

Title: US-09-905-083-32
Perfect score: 40
Sequence: 1 RLSSMVKKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	40	100.0	253	1 KLK7_HUMAN	P49662 homo sapien
2	33	82.5	372	1 COQ2_YEAST	P32378 saccharomyc
3	33	82.5	560	1 YM23_YEAST	P40215 saccharomyc
4	32	80.0	350	1 YUFN_BACSU	O05252 bacillus su
5	32	80.0	358	1 COQ2_SCHPO	Q10252 schizosacch
6	32	80.0	589	1 SYR_TREPA	O83803 treponema p
7	30	75.0	318	1 RCN2_RAT	Q62703 rattus norv
8	30	75.0	524	1 C90C_ARATH	Q9m066 arabidopsis
9	30	75.0	572	1 5NPD_LUTLO	Q9xz43 lutzomyia l
10	30	75.0	833	1 MBP1_YEAST	P39678 saccharomyc
11	29	72.5	253	1 T2D7_RAT	Q62880 rattus norv
12	29	72.5	571	1 CRT1_WAIZE	P49086 zea mays (m
13	29	72.5	582	1 CRT1_CAPAN	P80093 capsicum an
14	29	72.5	583	1 CRT1_LYCES	P28554 lycopersico
15	29	72.5	839	1 YNE1_YEAST	P53959 saccharomyc
16	28	70.0	41	1 Y010_TREPA	O83054 treponema p
17	28	70.0	109	1 YES7_YEAST	P40062 saccharomyc
18	28	70.0	175	1 PAP1_MOUSE	P35230 mus musculu
19	28	70.0	213	1 MAG2_SCHPO	O94468 schizosacch
20	28	70.0	264	1 RS3A_APLCA	P49395 aplysium cal
21	28	70.0	317	1 FADP_BACSU	P71019 bacillus su
22	28	70.0	448	1 FIBP_ADEP3	O83457 porcine ade
23	28	70.0	506	1 GTA_NPVAC	P41447 autographa
24	28	70.0	524	1 IMH3_YEAST	P50094 saccharomyc
25	28	70.0	553	1 YMB4_YEAST	Q04847 saccharomyc
26	28	70.0	625	1 HTPG_BACHD	Q9ke51 bacillus ha
27	28	70.0	690	1 PTA_MYCTU	P96254 mycobacteri
28	28	70.0	740	1 Y179_HUMAN	Q14684 homo sapien
29	28	70.0	851	1 YD72_SCHPO	Q10327 schizosacch
30	28	70.0	1021	1 BUB1_YEAST	P41695 saccharomyc
31	28	70.0	1165	1 LEPR_HUMAN	P48357 homo sapien
32	27	67.5	60	1 68MP_BOVIN	P14790 bos taurus
33	27	67.5	89	1 RS20_HELPJ	Q92mz1 helicobacte

ALIGNMENTS

RESULT 1

ID	KLK7_HUMAN	STANDARD;	PRT;	253 AA.
AC	P49662:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DE	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic enzyme) (hSCCE).			
GN	KLK7 OR PRSS6 OR SCCE.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.			
RC	TISSUE=Skin;			
RX	MEDLINE=94308225; PubMed=8034709;			
RA	Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A.,			
RA	Egelrud T.;			
RT	"Cloning, expression, and characterization of stratum corneum			
RT	chymotryptic enzyme. A skin-specific human serine proteinase.";			
RL	J. Biol. Chem. 269:19420-19426(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Keratinocytes;			
RA	Yousef G.M., Scorilas A., Diamandis E.P.;			
RT	"Molecular characterization, mapping and tissue expression of the			
RT	human stratum corneum chymotryptic enzyme gene.";			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20510030; PubMed=11054574;			
RA	Can L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,			
RA	Moss P., Paepier B., Wang K.;			
RT	"Sequencing and expression analysis of the serine protease gene			
RT	cluster located in chromosome 19q13 region.";			
RL	Gene 257:119-130(2000).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=95314630; PubMed=7794273;			
RA	Skytt A., Stroemqvist M., Egelrud T.;			
RT	"Primary substrate specificity of recombinant human stratum corneum			
RT	chymotryptic enzyme.";			
RL	Biochem. Biophys. Res. Commun. 211:586-589(1995).			
CC	-!- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE			
CC	STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS			
CC	SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID			
CC	RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SOCE			
CC	CLEAVES INSULIN B CHAIN AT 6-LEU-I-CYS-7, 16-TYR-I-LEU-17, 25-PHE-			
CC	I-TYR-26, AND 26-TYR-I-THR-27. COULD PLAY A ROLE IN THE ACTIVATION			
CC	OF PRECURSORS TO INFLAMMATORY CYTOKINES.			
CC	-!- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS			
CC	EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE			
CC	ALSO SEEN IN THE BRAIN AND KIDNEY.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			

P56027 helicobacte
P42038 cladosporiu
P42039 cladosporiu
P44123 haemophilus
P36540 escherichia
P78457 guillardia
P47347 mycoplasma
O20130 chlorella v
P96488 streptococc
Q14257 homo sapien
P57399 buchnera ap
P39849 pseudomonas

34 27 67.5 89 1 RS20_HELPJ
35 27 67.5 111 1 RLA3_CLAHE
36 27 67.5 112 1 RLA4_CLAHE
37 27 67.5 141 1 YB90_HAEIN
38 27 67.5 171 1 HSCB_ECOLI
39 27 67.5 183 1 PSAB_GUITH
40 27 67.5 222 1 Y101_MYCGE
41 27 67.5 250 1 YCX3_CHLVU
42 27 67.5 267 1 PROB_SPTTR
43 27 67.5 317 1 RCN2_HUMAN
44 27 67.5 319 1 TRXB_BUCAI
45 27 67.5 366 1 XYL8_PSEPU

```
CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L33404; AAC37551.1; -.
DR EMBL; AF166330; AAD49718.1; -.
DR EMBL; AF243527; AAG33360.1; -.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.300; -.
DR MIM; 604438; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal.
FT SIGNAL 1 22
FT PROPEP 23 29 ACTIVATION PEPTIDE.
FT CHAIN 30 253 KALLIKREIN 7.
FT ACT_SITE 70 70 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 36 137 BY SIMILARITY.
FT DISULFID 55 71 BY SIMILARITY.
FT DISULFID 137 239 BY SIMILARITY.
FT DISULFID 144 211 BY SIMILARITY.
FT DISULFID 176 190 BY SIMILARITY.
FT DISULFID 201 226 BY SIMILARITY.
FT CARBOHYD 246 246 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 253 AA; 27525 MW; 2D68B6B15A76A668 CRC64;

Query Match 100.08; Score 40; DB 1; Length 253;
Best Local Similarity 100.08; Pred. No. 0.13; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

Oy 1 RLSSMVKKV 9
|||
Db 123 RLSSMVKKV 131

RESULT 2
COQ2_YEAST
ID COQ2_YEAST STANDARD; PRT; 372 AA.
AC P32378;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Para-hydroxybenzoate-polyphenyltransferase, mitochondrial precursor
DE (EC 2.5.1.1) (PDB:polyphenyltransferase).
GN COQ2 OR YNR041C OR N3419.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92156158; PubMed=1740455;
RA Ashby M.N., Kutsunal S.Y., Ackerman S., Tzagoloff A., Edwards P.A.;
RT "COQ2 is a candidate for the structural gene encoding para-
RT hydroxybenzoate:polyphenyltransferase."
RL J. Biol. Chem. 267:4128-4136(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Pohl T.M.;

Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
-!- FUNCTION: CATALYZES THE PRENYLATION OF PARA-HYDROXYBENZOATE
WITH AN ALL-TRANS POLYPRENYL GROUP.
-!- PATHWAY: SECOND STEP IN COENZYME Q BIOSYNTHESIS.
-!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
inner membrane.
-!- SIMILARITY: BELONGS TO THE UBIA PRENYLTRANSFERASE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M81698; AAA34507.1; -.
DR EMBL; Z71656; CAA96321.1; -.
DR PIR; S20056; S20056.
DR SGD; S0005324; COQ2.
DR InterPro; IPR000537; UBIA.
DR Pfam; PF01040; UbIA; 1.
DR PROSITE; PS00943; UBIA; 1.
KW Isoprene biosynthesis; Transferase; Transit peptide; Mitochondrion;
Transmembrane.
FT TRANSIT 1 22
FT CHAIN 23 372
FT TRANSEM 92 112
FT TRANSEM 114 133
FT TRANSEM 170 191
FT TRANSEM 193 210
FT TRANSEM 229 249
FT TRANSEM 298 318
FT DOMAIN 29 49
FT DOMAIN 134 156
SQ SEQUENCE 372 AA; 41001 MW; 4D9738CE248B4AD6 CRC64;

Query Match 82.58; Score 33; DB 1; Length 372;
Best Local Similarity 77.88; Pred. No. 6.8; Indels 0; Gaps 0;
Matches 7; Conservative 1; Mismatches 1;

Oy 1 RLSSMVKKV 9
|||
Db 326 RLSSMVKKV 334

RESULT 3
YM23_YEAST
ID YM23_YEAST STANDARD; PRT; 560 AA.
AC P40215;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 62.8 kDa protein in RPS16A-TIF34 intergenic region.
GN YMR145C OR YG9375.14C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Badcock K., Churche C., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: STRONG, TO YEAST ROTENONE-INSENSITIVE NADH-UBIQUINONE
OX OXIDOREDUCTASE (ND11).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; Z47071; CAA87359.1; -.
 DR SGD; S0004753; YMR145C.
 DR InterPro; IPR001327; FAD_pyr_redox.
 DR Pfam; PF00070; pyr_redox; 1.
 KW Hypothetical protein; Oxidoreductase; Flavoprotein; FAD; NAD;
 KW Ubiquinone.
 SQ SEQUENCE 560 AA; 62774 MW; 10B1795E12E29C34 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 560;
 Best Local Similarity 77.8%; Pred. No. 11;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSWVKV 9
 ||:|||||
 Db 347 RLKTMVKV 355

RESULT 4
 YUFN_BACSU STANDARD; PRT; 350 AA.
 AC O05252;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 GN YUFN.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Oudega B., Koningshoeft A., Hilbert H.;
 RL Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
 (Potential).

CC -!- SIMILARITY: BELONGS TO THE BMP FAMILY OF LIPOPROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; Z39337; CAB07936.1; -.
 DR EMBL; Z39120; CAB15143.1; -.
 DR Subtilist; BG12349; yufn.
 DR InterPro; IPR003760; Bmp.
 DR Pfam; PF02608; Bmp; 1.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
 KW Hypothetical protein; Membrane; Lipoprotein; Signal;
 KW Complete proteome.

FT SIGNAL 1 13 PROBABLE.
 FT CHAIN 14 350 HYPOTHETICAL LIPOPROTEIN YUFN.
 FT LIPID 14 14 N-ACYL DIGLYCERIDE (PROBABLE).
 SQ SEQUENCE 350 AA; 37349 MW; 16D5176A52A99284 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 350;
 Best Local Similarity 87.5%; Pred. No. 11;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSWVKV 9
 |:|||||
 Db 269 LTSWVKV 276

RESULT 5
 COQ2_SCHPO STANDARD; PRT; 358 AA.
 AC Q10252;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Para-hydroxybenzoate--polyprenyltransferase, mitochondrial precursor
 (EC 2.5.1.-) (PDB:polyprenyltransferase) (P-hydroxybenzoate polyprenyl
 DE diphosphate transferase).
 GN COQ2 OR PP1 OR SPAC56F8.04C.

OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;

RN [1]
 RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
 RX MEDLINE=20545437; PubMed=11092853;
 RA Uchida N., Suzuki K., Saiki R., Kainou T., Tanaka K., Matsuda H.,
 RA Kawamukai M.;
 RT "Phenotypes of fission yeast defective in ubiquinone production due to
 RT disruption of the gene for p-hydroxybenzoate polyprenyl diphosphate
 RT transferase.";
 RL J. Bacteriol. 182:6933-6939(2000).

RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN=972;
 RA Pearson D., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
 RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.

CC -!- FUNCTION: CATALYZES THE PRENYLATION OF PARA-HYDROXYBENZOATE
 WITH AN ALL-TRANS POLYPRENYL GROUP (BY SIMILARITY).

CC -!- PATHWAY: SECOND STEP IN COENZYME Q BIOSYNTHESIS.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
 CC inner membrane.

CC -!- SIMILARITY: BELONGS TO THE UBIA PRENYLTRANSFERASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; AB053168; BAB20425.1; -.
 DR EMBL; Z69728; CAA93575.1; -.
 DR InterPro; IPR000537; Ubia.
 DR Pfam; PF01040; Ubia; 1.
 DR PROSITE; PS00943; UBIA; 1.

KW Isoprene biosynthesis; Transferase; Transit peptide; Mitochondrion;
 KW Transmembrane.

FT TRANSIT 1 ?
 FT CHAIN ? 358 ? MITOCHONDRION (POTENTIAL).
 PARA-HYDROXYBENZOATE--
 POLYPRENYLTRANSFERASE.

FT TRANSMEM 76 96 POTENTIAL.
 FT TRANSMEM 107 127 POTENTIAL.
 FT TRANSMEM 154 174 POTENTIAL.
 FT TRANSMEM 202 222 POTENTIAL.
 FT TRANSMEM 229 249 POTENTIAL.
 FT TRANSMEM 275 295 POTENTIAL.
 FT TRANSMEM 336 356 POTENTIAL.
 SQ SEQUENCE 358 AA; 39454 MW; AA5485411A0922F7 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 358;
 Best Local Similarity 77.8%; Pred. No. 11;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSWVKV 9
 |||||: ||
 Db 311 RLSSWVKV 319

of the cytochrome P-450 family that is required for the regulated polar elongation of leaf cells.";
Genes Dev. 12:2361-2391(1998).
[2]
SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G., Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N., Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M., Mueller M., Weichselgartner M., de Simone V., Obermaier B., Mache R., Schimidtheini T., Kreis M., Delsen M., Puigdomenech P., Watson M., Schmidtheini T., Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I., Vos P., Hohelsel J., Zimmermann W., Wedler H., Ridley P., Langham S.-A., McCullagh B., Bilham L., Robben J., Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F., Braeken M., Weijtjens I., Voet M., Bastiaens I., Aert R., Defoor E., Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M., Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W., Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P., Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H., Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S., Clark L., Doggett J., Hall S., Kay M., Lennard N., McTay K., Mayes R., Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S., Borkova D., Bloecker H., Scharfe M., Grimm M., Loechner T.-H., Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A., Neumann S., Argirou A., Vitale D., Liguori R., Piravandi E., Massenat O., Quigley F., Claubald G., Muendlein A., Felber R., Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S., Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E., Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A., Perez-Perez A., Fumelle B., Bent E., Johnson S., Tacon D., Jesse T., Reijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C., Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S., Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K., Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L., Sekhon M., Murray J., Sheet P., Cordes M., Abu-Theideh J., Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J., Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D., Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K., Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L., Nelson J., Speth J., Ryan E., Andrews S., Geisel C., Layman D., Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C., Antonoli B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C., Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R., Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S., Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A., Chen E., Marra M., Martienssen R., McCombie W.R.;
"Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.";
Nature 402:769-777(1999).
CC -!- FUNCTION: MIGHT BE INVOLVED IN THE BIOSYNTHESIS OF STEROIDS
CC IMPORTANT FOR THE POLAR ELONGATION OF CELLS DURING DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (Potential).
CC -!- TISSUE SPECIFICITY: UBQUITUOUS.
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AB008097; BAA37167.1;
DR EMBL; AL161589; CAB80304.1; ALT_INIT.
DR EMBL; Z99708; CAB16850.1; ALT_INIT.
DR EMBL; AL022141; CAA18139.1; ALT_SEQ.
DR InterPro; IPR001128; Cyt_p450.

DR Pfam; PF00067; p450; 1.
DR PRINTS; PR00385; P450.
DR PROSITE; PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme;
KW Endoplasmic reticulum; Multigene family.
FT TRANSMEM 4 24 POTENTIAL.
FT BINDING 463 463 HEME (BY SIMILARITY).
SQ SEQUENCE 524 AA; 59389 MW; 550578908BDDF272 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 524;
Best Local Similarity 77.8%; Pred. No. 44;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
DB 266 RLKMKVKV 274
II IIIII

RESULT 9
SNTD_LUTLO STANDARD; PRT; 572 AA.
ID SNTD_LUTLO STANDARD; PRT; 572 AA.
AC Q9XZ43;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 5'-nucleotidase precursor (EC 3.1.3.5).
GN 5NUC.
OS Lutzomyia longipalpis (Sand fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Psychodoidea;
OC Psychodidae; Lutzomyia.
OX NCBI_TaxID=7200;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JACOBINA; TISSUE=Salivary gland;
RA Charlab R., Valenzuela J.G., Rowton E.D., Ribeiro J.M.C.;
RT "Toward an understanding of the biochemical and pharmacological complexity of the saliva of a hematophagous sand fly, Lutzomyia longipalpis.";
RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP CHARACTERIZATION.
RC TISSUE=Salivary gland;
RX MEDLINE=20193550; PubMed=10727894;
RA Ribeiro J.M., Rowton E.D., Charlab R.;
RT "The salivary 5'-nucleotidase/phosphodiesterase of the hematophagous sand fly corrected, Lutzomyia longipalpis.";
RL Insect Biochem. Mol. Biol. 30:279-285(2000).
RN [3]
RP ERRATUM.
RA Ribeiro J.M., Rowton E.D., Charlab R.;
RL Insect Biochem. Mol. Biol. 30:609-609(2000).
CC -!- FUNCTION: DEGRADATION OF EXTERNAL UDP-GLUCOSE TO URIDINE MONOPHOSPHATE AND GLUCOSE-1-PHOSPHATE, WHICH CAN THEN BE USED BY THE CELL (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: UDP-SUGAR + H(2)O = UMP + SUGAR 1-PHOSPHATE.
CC -!- CATALYTIC ACTIVITY: A 5'-ribonucleotide + H(2)O = a ribonucleoside + phosphate.
CC -!- COFACTOR: ZINC (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE 5'-NUCLEOTIDASE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; AF132510; AAD32190.1;
DR HSP; P07024; ZUSH.
DR InterPro; IPR002224; 5_nucleotidase.
DR InterPro; IPR000934; Ser_thr_phosphatase.

```

DR Pfam: PF01009; 5_nucleotidase; 1.
DR Pfam: PF02872; 5_nucleotidaseC; 1.
DR PROSITE: PS00785; 5_NUCLEOTIDASE_1; 1.
DR PROSITE: PS00786; 5_NUCLEOTIDASE_2; 1.
KW Hydrolase; Glycoprotein; Signal; Zinc.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 572 5'-NUCLEOTIDASE.
FT CARBOHYD 82 82 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 454 454 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 490 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 572 AA; 63353 MW; 69A652338C04536D CRC64;

Query Match 75.0%; Score 30; DB 1; Length 572;
Best Local Similarity 75.0%; Pred. No. 48;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSWVKK 8
DB 70 RVSTMVKK 77

RESULT 10
MBPL_YEAST
ID MBPL_YEAST STANDARD; PRT; 833 AA.
AC P39678;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription factor MBP1 (MBF subunit P120).
GN MBP1 OR YDL056W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 372-387.
RC STRAIN=K1107;
RA Koch C., Moll T., Neuberg M., Ahorn H., Nasmyth K.;
RT "A role for the transcription factors Mbp1 and Swi4 in progression
RT from G1 to S phase.";
RL Science 261:1551-1557(1993).
RN [2]
RP SEQUENCE FROM N.A.
RA Bloecker H., Brandt P.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 460-833 FROM N.A.
RA Benton B.K., Plump S.D., Roos J., Lennarz W.J., Cross F.R.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 1-100.
RX MEDLINE=97446148; PubMed=9299332;
RA Taylor I.A., Treiber M.K., Olivi L., Smerdon S.J.;
RT "The x-ray structure of the DNA-binding domain from the Saccharomyces
RT cerevisiae cell-cycle transcription factor Mbp1 at 2.1-A
RT resolution.";
RL J. Mol. Biol. 272:1-8(1997).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.71 ANGSTROMS) OF 1-102.
RX MEDLINE=97238931; PubMed=9083114;
RA Xu R.M., Koch C., Liu Y., Horton J.R., Knapp D., Cheng X.;
RT "Crystal structure of the DNA-binding domain of Mbp1, a transcription
RT factor important in cell-cycle control of DNA synthesis.";
RL Structure 5:349-358(1997).
CC -!- FUNCTION: BINDS TO MCB ELEMENTS (MLU I CELL CYCLE BOX) FOUND IN
CC THE PROMOTER OF MOST DNA SYNTHESIS GENES. TRANSCRIPTIONAL
CC ACTIVATION BY MBF HAS AN IMPORTANT ROLE IN THE TRANSITION FROM G1
CC TO S PHASE. IT MAY HAVE A DUAL ROLE IN THAT IT BEHAVES AS AN
CC ACTIVATOR OF TRANSCRIPTION AT THE G1-S BOUNDARY AND AS A REPRESSOR
CC DURING OTHER STAGES OF THE CELL CYCLE.
CC -!- SUBUNIT: MBF CONTAINS SWI6 AND MBP1.
CC -!- SUBCELLULAR LOCATION: Nuclear.

```

```

CC -!- SIMILARITY: CONTAINS 2 ANK REPEATS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: X74158; CAA52271.1; -.
DR EMBL: Z74104; CAA98618.1; -.
DR EMBL: U19608; AAC49290.1; -.
DR PIR: A47528; A47528.
DR PIR: S37404; S37404.
DR PDB: 1MB1; 29-JUL-98.
DR PDB: 1BM8; 02-MAR-99.
DR TRANSFAC: T03480; -.
DR SGD: S0002214; MBP1.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR003163; Yeast_DNA_bind.
DR Pfam: PF00023; ank; 2.
DR Pfam: PF02292; Yeast_DNA_bind; 1.
DR SMART: SM00248; ANK; 2.
DR PROSITE: PS50088; ANK_REPEAT; 2.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
KW Transcription regulation; Activator; Nuclear protein; DNA-binding;
KW ANK repeat; Repeat; 3D-structure.
FT DNA_BIND 1 94
FT REPEAT 394 423 ANK 1.
FT REPEAT 512 541 ANK 2.
SQ SEQUENCE 833 AA; 93907 MW; BB7C35E29802BBD5 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 833;
Best Local Similarity 75.0%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSWVKK 8
DB 767 KLSLVKK 774

RESULT 11
T2D7_RAT
ID T2D7_RAT STANDARD; PRT; 253 AA.
AC Q62880;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcription initiation factor TFIID 31 kDa subunit (TAFII-31)
DE (TAFII-32) (TAFII32) (Neuronal cell death related gene in neuron -7)
DE (DN-7).
GN TAF9 OR TAF2G OR TAFII31.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97312553; PubMed=9168994;
RA Aoki T., Koike T., Nakano T., Shibahara K., Nishimura H., Kikuchi H.,
RA Honjo T.;
RT "Rat TAFII31 gene is induced upon programmed cell death in
RT differentiated PC12 cells deprived of NGF.";
RL Biochem. Biophys. Res. Commun. 234:230-234(1997).
CC -!- FUNCTION: TAFs ARE COMPONENTS OF THE TRANSCRIPTION FACTOR IID
CC (TFIID) COMPLEX THAT ARE ESSENTIAL FOR MEDIATING REGULATION OF RNA
CC POLYMERASE TRANSCRIPTION. TAFII31 IS A COACTIVATOR FOR THE P53
CC PROTEIN. ALSO INTERACTS WITH THE ACIDIC TRANSACTIVATOR VIRAL
CC PROTEIN 16 (VP16) AS WELL AS WITH THE GENERAL TRANSCRIPTION FACTOR
CC TFIIB (BY SIMILARITY).
CC -!- SUBUNIT: TFIID IS COMPOSED OF TATA BINDING PROTEIN (TBP) AND A
CC NUMBER OF TBP-ASSOCIATED FACTORS (TAFs) (BY SIMILARITY). TAFII31

```

```
CC AND P53 BIND TO EACH OTHER VIA AMINO ACID RESIDUES IN THE AMINO-
CC TERMINAL DOMAIN OF P53 THAT ARE ESSENTIAL FOR TRANSCRIPTION (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE TAF2G FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U40188; AAC53201.1; -.
CC InterPro: IPR000166; Histone_core.
CC InterPro: IPR003162; TFIID-31.
CC Pfam: PF02291; TFIID-31; 1.
CC ProDom: PD011023; TFIID-31; 1.
CC Transcription regulation; Nuclear protein.
CC DOMAIN 238 251 POLY-ASP.
CC SEQUENCE 253 AA; 27620 MW; 42045091A9B94378 CRC64;
CC -----
CC Query Match 72.5%; Score 29; DB 1; Length 253;
CC Best Local Similarity 75.0%; Pred. No. 37;
CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC -----
CC QY 1 RLSSMVKK 8
CC II I:III
CC DB 123 RLKSLVKK 130
CC -----
CC RESULT 12
CC CRTI_MAIZE
CC ID CRTI_MAIZE STANDARD; PRT; 571 AA.
CC AC P49086; Q41849;
CC DT 01-FEB-1996 (Rel. 33, Created)
CC DT 01-FEB-1996 (Rel. 33, Last sequence update)
CC DT 01-MAR-2002 (Rel. 41, Last annotation update)
CC DE Phytoene dehydrogenase, chloroplast precursor (EC 1.14.99.-)
CC (Phytoene desaturase).
CC PDS1 OR PDS.
CC OS Zea mays (Maize).
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade;
CC Panicoideae; Andropogoneae; Zea.
CC NCBI_TaxID=4577;
CC [1]
CC SEQUENCE FROM N.A.
CC MEDLINE=96178866; PubMed=8616251;
CC Li Z., Matthews P.D., Burr B., Wurtzel E.T.;
CC "Cloning and characterization of a maize cDNA encoding phytoene
CC desaturase, an enzyme of the carotenoid biosynthetic pathway.";
CC Plant Mol. Biol. 30:269-279(1996).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=CV. FUNKF; TISSUE=Leaf;
CC MEDLINE=95357435; PubMed=7630964;
CC Habie W.E., Olshi K.K.;
CC "Maize phytoene desaturase maps near the viviparous5 locus.";
CC Plant Physiol. 108:1329-1330(1995).
CC -!- FUNCTION: THIS ENZYME CONVERTS PHYTOENE INTO ZETA-CAROTENE VIA THE
CC INTERMEDIARY OF PHYTOFLUENE BY THE SYMMETRICAL INTRODUCTION OF TWO
CC DOUBLE BONDS AT THE C-11 AND C-11' POSITIONS OF PHYTOENE.
CC -!- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -!- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -!- SIMILARITY: BELONGS TO THE PHYTOENE DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U37285; AAC12846.1; -.
CC EMBL: L39266; AAA99519.1; -.
CC MaizedB: 84977; -.
CC Mendel: 13463; Zeama:Pds1;13463.
CC InterPro: IPR002937; Amino_oxidase.
CC InterPro: IPR000205; NAD_binding.
CC Pfam: PF01593; Amino_oxidase; 1.
CC Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;
CC Chloroplast; Transit peptide.
CC TRANSIT 1 96
CC CHAIN 97 571 CHLOROPLAST (POTENTIAL).
CC PHYTOENE DEHYDROGENASE.
CC FT CONFLICT 61 63 VVC -> LSA (IN REF. 2).
CC FT CONFLICT 68 68 R -> S (IN REF. 2).
CC FT CONFLICT 555 555 A -> T (IN REF. 2).
CC SQ SEQUENCE 571 AA; 64115 MW; FAEL19C7EFBE799A CRC64;
CC -----
CC Query Match 72.5%; Score 29; DB 1; Length 571;
CC Best Local Similarity 55.6%; Pred. No. 80;
CC Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
CC -----
CC QY 1 RLSSMVKV 9
CC II I:III
CC DB 329 RLNSRIKKI 337
CC -----
CC RESULT 13
CC CRTI_CAPAN
CC ID CRTI_CAPAN STANDARD; PRT; 582 AA.
CC AC P80093;
CC DT 01-DEC-1992 (Rel. 24, Created)
CC DT 01-DEC-1992 (Rel. 24, Last sequence update)
CC DT 01-MAR-2002 (Rel. 41, Last annotation update)
CC DE Phytoene dehydrogenase, chloroplast precursor (EC 1.14.99.-)
CC (Phytoene desaturase).
CC PDS.
CC OS Capsicum annuum (Bell pepper).
CC OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
CC Asteridae; euasterids I; Solanales; Solanaceae; Capsicum.
CC NCBI_TaxID=4072;
CC [1]
CC SEQUENCE FROM N.A.
CC STRAIN=CV. LAMUYO;
CC MEDLINE=93011154; PubMed=1396714;
CC Huguency P., Roemer S., Kuntz M., Camara B.;
CC "Characterization and molecular cloning of a flavoprotein catalyzing
CC the synthesis of phytofluene and zeta-carotene in Capsicum
CC chromoplasts.";
CC Eur. J. Biochem. 209:399-407(1992).
CC -!- FUNCTION: THIS ENZYME CONVERTS PHYTOENE INTO ZETA-CAROTENE VIA THE
CC INTERMEDIARY OF PHYTOFLUENE BY THE SYMMETRICAL INTRODUCTION OF TWO
CC DOUBLE BONDS AT THE C-11 AND C-11' POSITIONS OF PHYTOENE.
CC -!- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -!- ENZYME REGULATION: INHIBITED BY THE HERBICIDES METFLURAZON,
CC DIFUNONE, FLURIDONE AND DIFLUFENICAN.
CC -!- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -!- DEVELOPMENTAL STAGE: RIPENING FRUIT.
CC -!- SIMILARITY: BELONGS TO THE PHYTOENE DEHYDROGENASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X68058; CAA48195.1; -.
CC
```



```

DR PIR: S29314; S29314.
DR InterPro: IPR002937; Amino_Oxidase.
DR InterPro: IPR00205; NAD_binding.
DR Pfam: PF01593; Amino_Oxidase; 1.
KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;
FT TRANSIT 1 110 CHLOROPLAST (POTENTIAL).
FT CHAIN 111 582 PHYTOENE DEHYDROGENASE.
FT NP_BIND 117 133 FAD (ADP PART) (POTENTIAL).
SQ SEQUENCE 582 AA; 65061 MW; 011E6E7DCAFB3DB5 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 582;
Best Local Similarity 55.6%; Pred. No. 82;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
DB 343 RLNSRIKKI 351

RESULT 14
CRTL_LYCES STANDARD; PRT; 583 AA.
AC P28554;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Phytoene dehydrogenase, chloroplast precursor (EC 1.14.99.-)
DE (Phytoene desaturase).
GN PDS.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92279247; PubMed=1594600;
RA Pecker I., Chamovitz D., Hirschberg J.;
RT "A single polypeptide catalyzing the conversion of phytoene to zeta-
RT carotene is transcriptionally regulated during tomato fruit
RT ripening.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:4962-4966(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=CV, AILSA CRAIG; TISSUE=Fruit;
RX MEDLINE=93250433; PubMed=8485401;
RA Bartley G.B., Scolnik P.A., Giuliano G.;
RT "Regulation of carotenoid biosynthesis during tomato development.";
RL Plant Cell 5:379-387(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94169297; PubMed=8123786;
RA Mann V., Pecker I., Hirschberg J.;
RT "Cloning and characterization of the gene for phytoene desaturase
RT (Pds) from tomato (Lycopersicon esculentum).";
RL Plant Mol. Biol. 24:429-434(1994).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=CV, AILSA CRAIG;
RX MEDLINE=95083762; PubMed=7991692;
RA Aracri B., Bartley G.E., Scolnik P.A., Giuliano G.;
RT "Sequence of the phytoene desaturase locus of tomato.";
RL Plant Physiol. 105:789-789(1994).
CC -!- FUNCTION: THIS ENZYME CONVERTS PHYTOENE INTO ZETA-CAROTENE VIA THE
CC INTERMEDIARY OF PHYTOFLUENE BY THE SYMMETRICAL INTRODUCTION OF TWO
CC DOUBLE BONDS AT THE C-11 AND C-11' POSITIONS OF PHYTOENE.
CC -!- COFACTOR: NAD, NADP, OR FAD (PROBABLE).
CC -!- PATHWAY: CAROTENOID BIOSYNTHESIS.
CC -!- SUBCELLULAR LOCATION: Chloroplast; chromoplast.
CC -!- DEVELOPMENTAL STAGE: RIPENING FRUIT.
CC -!- SIMILARITY: BELONGS TO THE PHYTOENE DEHYDROGENASE FAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: X59948; CAA42573.1; -.
DR EMBL: M88683; AAA68865.1; -.
DR EMBL: X71023; -. NOT_ANNOTATED_CDS.
DR EMBL: X78271; CAA55078.1; -.
DR PIR: S21502; S21502.
DR PIR: A45381; A45381.
DR PIR: S42544; S42544.
DR InterPro: IPR002937; Amino_Oxidase.
DR InterPro: IPR00205; NAD_binding.
DR Pfam: PF01593; Amino_Oxidase; 1.
KW Carotenoid biosynthesis; Oxidoreductase; NAD; Flavoprotein; FAD;
FT TRANSIT 1 111 CHLOROPLAST (POTENTIAL).
FT CHAIN 112 583 PHYTOENE DEHYDROGENASE.
FT NP_BIND 118 134 FAD (ADP PART) (POTENTIAL).
FT CONFLICT 508 508 T -> A (IN REF. 2).
SQ SEQUENCE 583 AA; 64884 MW; 9EA0DF71A2FA3A44 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 583;
Best Local Similarity 55.6%; Pred. No. 82;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
DB 344 RLNSRIKKI 352

RESULT 15
YNEI_YEAST STANDARD; PRT; 839 AA.
AC P53959;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 97.0 kDa protein in YIP3-TPC5 intergenic region.
GN YNL041C OR N2675.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA Duesterhoeft A., Floeth M., Fritz C., Heuss-Neitzel D.,
RA Hilbert H., Moestl D.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: SOME, S.POMBE SPBC776.10C.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: Z71317; CAA95908.1; -.
DR SGD: S0004986; YNL041C.
KW Hypothetical protein.
SQ SEQUENCE 839 AA; 96975 MW; 6484A40F999AD787 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 839;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

```



```

Db 227 RLSSSVEKI 235
|||||:|:|:
ID Y010_TREPA STANDARD; PRT; 41 AA.
AC O83054;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein TP0010.
GN TP0010.
OS Treponema pallidum.
OC Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX NCBI_TaxID=160;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NICHOLS;
RX MEDLINE=98332770; PubMed=9665876;
RA Fraser C.M., Norris S.J., Weinstein G.M., White O., Sutton G.G.,
RA Dodson R., Winn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA Sodergren E., Hardham J.M., McLeod M.P., Salzberg S., Peterson J.,
RA Khalak H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA Hatch B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA Venter J.C.;
RT "Complete genome sequence of Treponema pallidum, the syphilis
RT spirochete.";
RL Science 281:375-388(1998).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE001186; AAC65011.1; -.
DR TIGR; TP0010; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 41 AA; 4471 MW; 08178A6F9D5D83FB CRC64;

Query Match 70.0%; Score 28; DB 1; Length 41;
Best Local Similarity 55.6%; Pred. No. 11;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSSVKVK 9
:|:|:|:|:
Db 26 KLSEVVKI 34

RESULT 17
YES7_YEAST
ID YES7_YEAST STANDARD; PRT; 109 AA.
AC P40062;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Hypothetical 12.7 kDa protein in RAD51-UBP9 intergenic region.
GN YER097W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Lashkari D., Lew H., Lin D.,

```

```

RA Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Velton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U18839; AAB64652.1; -.
DR SGD; S0000899; YER097W.
KW Hypothetical protein.
SQ SEQUENCE 109 AA; 12728 MW; C675176AFC5525D8 CRC64;

Query Match 70.0%; Score 28; DB 1; Length 109;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSWVKV 9
:|:|:|:|:
Db 79 LSSILKKV 86

RESULT 18
PAP1_MOUSE
ID PAP1_MOUSE STANDARD; PRT; 175 AA.
AC P35230;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Pancreatitis-associated protein 1 precursor (REG III-beta).
GN PAP1 OR PAP OR REG3B.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas, and Small intestine;
RX MEDLINE=93176807; PubMed=7679928;
RA Itoh T., Teraoka H.;
RT "Cloning and tissue-specific expression of cDNAs for the human and
RL mouse homologues of rat pancreatitis-associated protein (PAP).";
RL Biochim. Biophys. Acta 1172:184-186(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=97208868; PubMed=9055810;
RA Narushima Y., Unno M., Nakagawara K.-I., Mori M., Miyashita H.,
RA Suzuki Y., Noguchi N., Takasawa S., Kumagai T., Yonekura H.,
RA Okamoto H.;
RT "Structure, chromosomal localization and expression of mouse genes
RT encoding type III Reg, RegIII alpha, RegIII beta, RegIII gamma.";
RL Gene 185:159-168(1997).
CC -!- FUNCTION: MIGHT BE A STRESS PROTEIN INVOLVED IN THE CONTROL OF
CC BACTERIAL PROLIFERATION.
CC -!- SUBCELLULAR LOCATION: SECRETED. FOUND IN THE APICAL REGION OF
CC PANCREATIC ACINAR CELLS (BY SIMILARITY).
CC -!- TISSUE SPECIFICITY: CONSTITUTIVELY EXPRESSED IN THE SMALL
CC INTESTINE, MODERATELY IN COLON AND AT AN EXTREMELY LOW LEVEL IN
CC HEALTHY PANCREAS.
CC -!- INDUCTION: APPEARS IN PANCREATIC JUICE AFTER INDUCTION OF
CC PANCREATIC INFLAMMATION.
CC -!- DISEASE: OVEREXPRESSED DURING THE ACUTE PHASE OF PANCREATITIS.
CC -!- SIMILARITY: TO OTHER MEMBERS OF THE C-TYPE LECTIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL; D13509; BAA02727.1; -
DR EMBL; D63359; BAA18928.1; -
DR EMBL; D63360; BAA18929.1; -
DR PIR; S29822; S29822.
DR HSSP; P05451; IQDD.
DR MGD; MGI:97478; Pep.
DR InterPro; IPR001304; lectin_c.
DR InterPro; IPR003990; pancreatitis_assoc.
DR Pfam; PF00059; lectin_c; 1.
DR PRINTS; PR01504; PNCREATISAP.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE-LECTIN_2; 1.
KW Signal; Lectin; Inflammatory response; Acute phase; Multigene family.
FT SIGNAL 1 26
FT CHAIN 27 175 PNCREATITIS-ASSOCIATED PROTEIN 1.
FT DOMAIN 38 173 C-TYPE LECTIN (LONG FORM).
FT DISULFID 40 51 BY SIMILARITY.
FT DISULFID 68 171 BY SIMILARITY.
FT DISULFID 146 163 BY SIMILARITY.
FT SEQUENCE 175 AA; 19476 MW; 44B3101171E79775 CRC64;
Query Match 70.0%; Score 28; DB 1; Length 175;
Best Local Similarity 85.7%; Pred. No. 43;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LSSMVKK 8
Db 88 LSSMVKR 94
|||||

RESULT 19
MAG2_SCHPO
ID MAG2_SCHPO STANDARD; PRT; 213 AA.
AC O94468;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-3-methyladenine glycosylase (EC 3.2.21) (3-methyladenine DNA glycosidase) (3MEA DNA glycosylase).
GN SPBC2367.14.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Xiang Z., Aves S., Wood V., Rajandream M.A., Barrell B.G.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HYDROLYSIS OF THE DEOXYRIBOSE N-GLYCOSIDIC BOND TO EXCISE 3-METHYLADENE OR 7-METHYLADENE FROM THE DAMAGED DNA POLYMER FORMED BY ALKYLATION LESIONS (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: Hydrolysis of alkylated DNA, releasing 3-methyladenine, 3-methylguanine, 7-methylguanine, and 7-methyladenine.
CC -!- SIMILARITY: BELONGS TO THE ALKYLBASE DNA GLYCOSIDASES ALKA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; AL035065; CAA22627.1; -

DR InterPro; IPR000035; Alkylbase_DNA_glycos.
DR InterPro; IPR003265; Endo_3c.
DR Pfam; PF00730; HHH-GPD; 1.
DR SMART; SM00478; ENDO3c; 1.
DR PROSITE; PS00516; ALKYLBASE_DNA_GLYCOS; FALSE_NEG.
KW DNA repair; Hydrolase; Multigene family.
SQ SEQUENCE 213 AA; 24292 MW; 21E7E34A181D80EE CRC64;
Query Match 70.0%; Score 28; DB 1; Length 213;
Best Local Similarity 85.7%; Pred. No. 52;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 SSMVKKV 9
Db 22 SSMVKKV 28
|||||

RESULT 20
RS3A_APLCA
ID RS3A_APLCA STANDARD; PRT; 264 AA.
AC P49395;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE 40S ribosomal protein S3A (lysine-rich protein KRP-A).
GN RPS3A OR KRP-A.
OS Aplysia californica (California sea hare).
OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia; Anasplidea;
OC Aplysiidae; Aplysia.
OX NCBI_TaxID=6500;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuron;
RX MEDLINE=94192693; PubMed=8143753;
RA Auclair D., Lang B.L., Desgroseillers L., Forest P.;
RT "Analysis of genes encoding highly conserved lysine-rich proteins in Aplysia californica and Saccharomyces cerevisiae.";
RL Eur. J. Biochem. 220:997-1003(1994).
CC -!- SIMILARITY: BELONGS TO THE S3AE FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X68555; CAA48558.1; -
DR InterPro; IPR001593; Ribosomal_S3AE.
DR Pfam; PF01015; Ribosomal_S3AE; 1.
DR ProDom; PD003035; Ribosomal_S3AE; 1.
DR PROSITE; PS01191; RIBOSOMAL_S3AE; 1.
KW Ribosomal protein.
FT INIT_MET 0
FT SEQUENCE 264 AA; 29528 MW; BFF8DB3A1E4B4F12 CRC64;
Query Match 70.0%; Score 28; DB 1; Length 264;
Best Local Similarity 75.0%; Pred. No. 64;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
Db 108 RLSSMVKK 115
|||||

RESULT 21
FABD_BACSU
ID FABD_BACSU STANDARD; PRT; 317 AA.
AC P71019; O34463;
DT 01-NOV-1997 (Rel. 35, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

DE Malonyl CoA-acyl carrier protein transacylase (EC 2.3.1.39) (MCT).
GN FAD.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=96326321; PubMed=8759840;
RA Morbidoni H.R., de Mendoza D., Cronan J.E. Jr.;
RT "Bacillus subtilis acyl carrier protein is encoded in a cluster of
RT lipid biosynthesis genes.";
RL J. Bacteriol. 178:4794-4800(1996).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=168;
RX MEDLINE=98195738; PubMed=9534248;
RA Foulger D., Errington J.;
RT "A 28 kbp segment from the spoVM region of the Bacillus subtilis 168
RT genome.";
RL Microbiology 144:801-805(1998).
CC -!- CATALYTIC ACTIVITY: Malonyl-CoA + [acyl-carrier protein] = CoA +
CC malonyl-[acyl-carrier protein].
CC -!- PATHWAY: FATTY ACID BIOSYNTHESIS.
CC -!- SIMILARITY: SIGNIFICANT, TO SEVERAL OTHER PROTEINS WITH
CC TRANSACYLASE ACTIVITY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U59433; AAC44306.1; -
DR EMBL; Y13937; CAA74249.1; -
DR EMBL; Z99112; CAB13463.1; -
DR HSSP; P25715; LMLA.
DR Subtilist; BG11836; fabd.
DR InterPro; IPR001227; Acyltransf_domain.
DR Pfam; PF00698; Acyl_transf; 1.
KW Fatty acid biosynthesis; Complete proteome.
FT ACT_SITE 91 91 BY SIMILARITY.
FT ACT_SITE 201 201 BY SIMILARITY.
FT CONFLICT 124 131 VPAGEGAM -> GCRLAKEQW (IN REF. 1).
SQ SEQUENCE 317 AA; 34035 MW; 191AE828B1C91D7F CRC64;

Query Match 70.0%; Score 28; DB 1; Length 317;
Best Local Similarity 75.0%; Pred. No. 76;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LSSMVKKV 9
DB 281 LSGLVKKV 288

RESULT 22
ID FIBP_ADEP3 STANDARD; PRT; 448 AA.
AC Q83457; Q9VTR2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Fiber protein.
GN PIV.
OS Porcine adenovirus type 3 (PAV-3).
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Mastadenovirus.
OX NCBI_TaxID=35265;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=6618;

```

```

RX MEDLINE=95351046; PubMed=7625129;
RA Reddy P.S., Nagy E., Derbyshire J.B.;
RT "Sequence analysis of putative pVIII, E3 and fibre regions of porcine
RT adenovirus type 3.";
RL Virus Res. 36:97-106(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=6618;
RX MEDLINE=99058191; PubMed=9837805;
RA Reddy P.S., Idamakanti N., Song J.Y., Lee J.B., Hyun B.H., Park J.H.,
RA Cha S.H., Bae Y.T., Tikoo S.K., Babiuk L.A.;
RT "Nucleotide sequence and transcription map of porcine adenovirus type
RT 3.";
RL Virology 251:414-426(1998).
RN [3]
RP SEQUENCE FROM N.A.
RX STRAIN=6618 / IAF;
RA Larocque D., Malenfant F., Massie B., Dea S.;
RT "Porcine adenovirus serotype 3, complete genome.";
RT Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: RECOGNIZES THE CELL RECEPTOR; SERVES AS THE LIGAND
CC -!- BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR.
CC -!- SUBUNIT: HOMOTRIMER (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF083132; AAC99449.1; -
DR EMBL; AJ237815; CAB41035.1; -
DR EMBL; AB026117; BAA76973.1; -
DR InterPro; IPR000939; Adeno_fiber2.
DR InterPro; IPR000978; Adeno_fiber_knob.
DR Pfam; PF00541; adeno_fiber; 1.
DR Pfam; PF00608; adeno_fiber2; 4.
DR PRINTS; PR00307; ADENOVSFIBRE.
KW Fiber protein.
SQ SEQUENCE 448 AA; 47846 MW; 8B929F4D16A0FA3 CRC64;

Query Match 70.0%; Score 28; DB 1; Length 448;
Best Local Similarity 75.0%; Pred. No. 11e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LSSMVKKV 9
DB 307 LTSMTKKV 314

RESULT 23
ID GTA_NPVAC STANDARD; PRT; 506 AA.
AC P41447;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Probable global transactivator.
GN GTA.
OS Autographa californica nuclear polyhedrosis virus (ACMNPV).
OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;
OC Nucleopolyhedrovirus.
OX NCBI_TaxID=46015;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C6;
RX MEDLINE=94303173; PubMed=8030224;
RA Ayres M.D., Howard S.C., Kuzio J., Lopez-Ferber M., Possee R.D.;
RT "The complete DNA sequence of Autographa californica nuclear
RT polyhedrosis virus.";

```

```
RL Virology 202:586-605(1994).
CC -!- SIMILARITY: BELONGS TO THE SNF2/RAD54 HELICASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L22858; AAA66672.1;
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR InterPro; IPR003330; SNF2_N.
DR Pfam; PF00271; helicase_C; 1.
DR Pfam; PF00176; SNF2_N; 1.
DR SMART; SM00487; DEXDC; 1.
DR SMART; SM00490; HELICC; 1.
DR KW Helicase; ATP-binding.
FT NP_BIND 58 64 ATP (BY SIMILARITY).
FT SITE 157 160 DEAD BOX.
FT SEQUENCE 506 AA; 59058 MW; 84E0CC24A85984AE CRC64;
SQ
Query Match 70.0%; Score 28; DB 1; Length 506;
Best Local Similarity 44.4%; Pred. No. 1.2e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;
QY 1 RLSSWKKV 9
Db 226 RKSIKKI 234
RESULT 24
IMH3_YEAST
ID IMH3_YEAST STANDARD; PRT; 524 AA.
AC P50094;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Probable inosine-5'-monophosphate dehydrogenase (EC 1.1.1.205) (IMP
DE dehydrogenase) (IMPDH) (IMPD).
GN YML056C OR YM958.06C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Devlin K., Church C., Barrell B.G., Rajandream M.A.;
RL Submitted (NOV-1994) to the EMBL/GenBank/DDBJ databases.
CC -!- CATALYTIC ACTIVITY: Inosine 5'-phosphate + NADH.
CC xanthosine 5'-phosphate + NADH.
CC -!- PATHWAY: FIRST REACTION UNIQUE TO GMP BIOSYNTHESIS.
CC -!- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC IMPDH AND TO
CC GMP REDUCTASE.
CC -!- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 246729; CAA86719.1;
DR HSSP; P12268; 1B30.
DR SGD; S0004520; YML056C.
DR InterPro; IPR000844; CBS.
DR InterPro; IPR003009; FMN_enzyme.
DR InterPro; IPR001093; IMP_DH_GMP_RED.
```

```
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF00478; IMPDH_C; 1.
DR Pfam; PF01574; IMPDH_N; 1.
DR SMART; SM00116; CBS; 2.
DR PROSITE; PS00487; IMP_DH_GMP_RED; 1.
DR KW Hypothetical protein; Oxidoreductase; NAD; GMP biosynthesis;
KW purine biosynthesis; Multigene family; Repeat; CBS domain.
FT DOMAIN 120 175 CBS 1.
FT DOMAIN 183 235 CBS 2.
FT BINDING 336 336 IMP (POTENTIAL).
SQ SEQUENCE 524 AA; 56394 MW; A73D1E4EFE8AEAD9 CRC64;
Query Match 70.0%; Score 28; DB 1; Length 524;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 SSMWKKV 9
Db 109 ASMWKKV 115
RESULT 25
YM8A_YEAST
ID YM8A_YEAST STANDARD; PRT; 553 AA.
AC Q04847;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Hypothetical 64.4 kDa protein in PET111-TIF11 intergenic region.
GN YMR258C OR YM9920.12C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Hunt S., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-1995) to the EMBL/GenBank/DDBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; 248639; CAA88585.1;
DR SGD; S0004871; YMR258C.
KW Hypothetical protein.
SQ SEQUENCE 553 AA; 64405 MW; 2F764964A6C23EC2 CRC64;
Query Match 70.0%; Score 28; DB 1; Length 553;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 RLSSWKKV 9
Db 155 RLSSLKKRV 163
RESULT 26
HTPG_BACHD
ID HTPG_BACHD STANDARD; PRT; 625 AA.
AC Q9KE51;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Chaperrone protein htpg (Heat shock protein htpg) (High temperature
DE protein G).
GN HTPG OR BH1007.
OS Bacillus halodurans.
```

OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 RT halodurans and genomic sequence comparison with Bacillus subtilis.";
 RL Nucleic Acids Res. 28:4317-4331(2000).
 CC -!- FUNCTION: Molecular chaperone. Has ATPase activity (By
 CC similarity).
 CC -!- SUBUNIT: Homodimer (By similarity).
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -!- SIMILARITY: BELONGS TO THE HEAT SHOCK PROTEIN 90 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AP001510; BAB04726.1; -;
 DR HSP: P07900; IYER.
 DR InterPro: IPR003594; HATPase_c.
 DR InterPro: IPR001404; HSP90.
 DR Pfam: PF02518; HATPase_c; 1.
 DR Pfam: PF00183; HSP90; 2.
 DR PRINTS: PR00775; HEATSHOCK90.
 DR SMART: SM00387; HATPase_c; 1.
 DR PROSITE: PS00298; HSP90; 1.
 DR Chapterone; ATP-binding; Heat shock; Complete proteome.
 KW DOMAIN 1 341 A; SUBSTRATE-BINDING (BY SIMILARITY).
 FT DOMAIN 342 525 B (BY SIMILARITY).
 FT DOMAIN 552 625 C.
 SQ SEQUENCE 625 AA; 72342 MW; FE784C96267DDBCF CRC64;
 Query Match 70.0%; Score 28; DB 1; Length 625;
 Best Local Similarity 62.5%; Pred. No. 1.5e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 RLSSMYKK 8
 Db 190 RLTSIIKK 197
 ||:||||
 RESULT 27
 PTA_MYCTU STANDARD; PRT; 690 AA.
 ID PTA_MYCTU
 AC P96254;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Phosphate acetyltransferase (EC 2.3.1.8) (Phosphotransacetylase).
 GN PTA OR RV0408 OR MT0421 OR MTCY22G10.04.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,

RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence.";
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Wadman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 CC -!- CATALYTIC ACTIVITY: Acetyl-CoA + phosphate = CoA + acetyl
 CC phosphate.
 CC -!- PATHWAY: LAST OF TWO STEPS IN THE CONVERSION OF ACETATE TO
 CC ACETYL-COA.
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
 CC -!- SIMILARITY: IN THE N-TERMINAL SECTION; CONTAINS A DOMAIN WITH SOME
 CC SIMILARITIES WITH COBQ/BIOD.
 CC -!- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE PHOSPHATE
 CC ACETYLTRANSFERASE AND BUTYRYLTRANSFERASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z84724; CAB06578.1; -;
 DR EMBL: AE006946; AAK44645.1; -;
 DR TIGR: MT0421; -;
 DR Tuberculist: Rv0408; -;
 DR InterPro: IPR002505; PTA_PTB.
 DR Pfam: PF01515; PTA_PTB; 1.
 KW Transferase; Acyltransferase; Complete proteome.
 FT DOMAIN 365 690 PHOSPHATE ACETYLTRANSFERASE.
 SQ SEQUENCE 690 AA; 72948 MW; C01C412AF2810CCE CRC64;
 Query Match 70.0%; Score 28; DB 1; Length 690;
 Best Local Similarity 66.7%; Pred. No. 1.6e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 RLSSMYKKV 9
 Db 28 RLTAHVAKV 36
 ||:|||||
 RESULT 28
 Y179_HUMAN STANDARD; PRT; 740 AA.
 ID Y179_HUMAN
 AC Q14684;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical protein KIAA0179.
 GN KIAA0179.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Bone marrow;
 RX MEDLINE=96281124; PubMed=8724849;
 RA Nagase T., Seki N., Ishikawa K.-I., Tanaka A., Nomura N.;
 RT "Prediction of the coding sequences of unidentified human genes. V.

RT The coding sequences of 40 new genes (KIAA0161-KIAA0200) deduced by
RL analysis of cDNA clones from human cell line KG-1.¹;
RT DNA Res. 3:17-24(1996).
CC 1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC 1- SIMILARITY: BELONGS TO THE NNP-1 FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D80001; BA011496.1; ALT INIT
KW Hypothetical protein; Nuclear protein.
FT DOMAIN 384 388
SQ SEQUENCE 740 AA; 82159 MW; 2C76AD1FEB0FAC73 CRC64;

Query Match 70.0%; Score 28; DB 1; Length 740;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
DB 305 RLSKLIKK 312
||| :||

RESULT 29
YD72_SCHPO STANDARD; PRT; 851 AA.
AC Q10327;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE Hypothetical 97.1 kDa protein C32A11.02C in chromosome I.
GN SPAC32A11.02C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Skelton J.; Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DDBB databases.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z69796; CAA93699.1; -
KW Hypothetical protein.
FT DOMAIN 851 851
SQ SEQUENCE 851 AA; 97078 MW; 7E28BA445345DA21 CRC64;

Query Match 70.0%; Score 28; DB 1; Length 851;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
DB 369 QLSSEVKKI 377
:|||||

RESULT 30
BUB1_YEAST STANDARD; PRT; 1021 AA.
ID BUB1_YEAST
AC P41695;

DT 01-NOV-1995 (Rel. 32, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Checkpoint serine/threonine-protein kinase.BUB1 (EC 2.7.1.-).
DE BUB1 OR YGRI88C OR G7542.
GN Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=95059057; PubMed=7969164;
RA Roberts B.T., Farr K.A., Hoyt M.A.;
RT "The Saccharomyces cerevisiae checkpoint gene BUB1 encodes a novel
RT protein kinase.";
RL Mol. Cell. Biol. 14:8282-8291(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97279231; PubMed=9133739;
RA Arroyo J., Garcia-Gonzalez M., Garcia-Saez M.I., Sanchez-Perez M.,
RA Nombela C.;
RT "DNA sequence analysis of a 23,002 bp DNA fragment of the right arm
RT of Saccharomyces cerevisiae chromosome VII.";
RL Yeast 13:357-363(1997).
CC 1- FUNCTION: INVOLVED IN CELL CYCLE CHECKPOINT ENFORCEMENT. CATALYZES
CC THE PHOSPHORYLATION OF BUB3 AND ITS AUTOPHOSPHORYLATION.
CC 1- SUBCELLULAR LOCATION: Nuclear.
CC 1- PTM: AUTOPHOSPHORYLATED.
CC 1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
CC BUB1 SUBFAMILY
CC 1- SIMILARITY: CONTAINS 1 CD1 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L32027; AAA64894.1; -
DR EMBL; Z72973; CAA97214.1; -
DR EMBL; X99074; CAA67524.1; -
DR SGD; S0003420; BUB1.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 2.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
KW Transferrase; Serine/threonine-protein kinase; ATP-binding;
KW Cell cycle; Phosphorylation; Nuclear protein.
FT DOMAIN 55 304
FT CD1.
FT DOMAIN 705 1021
FT NP_BIND 711 719
FT BINDING 733 733
FT ACT_SITE 833 833
FT MUTAGEN 733 733
FT CONFLICT 531 531
SQ SEQUENCE 1021 AA; 117868 MW; 6D76FC980775D3F9 CRC64;

Query Match 70.0%; Score 28; DB 1; Length 1021;
Best Local Similarity 44.4%; Pred. No. 2.3e+02;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
DB 675 KWSLLKKI 683
:|||||

RESULT 31

```

LEPR_HUMAN
ID LEPR_HUMAN STANDARD; PRT; 1165 AA.
AC P48357;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Leptin receptor precursor (LEP-R) (OB receptor) (OB-R).
GN LEPR OR OBR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=96128129; PubMed=8548812;
RA Tartaglia L.A., Dembski M., Wang X., Deng N., Culpepper J.,
RA Devos R., Richards G.J., Campfield L.A., Clark F.T., Deeds J.,
RA Muir C., Sanker S., Morlarty A., Moore K.J., Smutko J.S.,
RA Mays G.G., Woolf E.A., Monroe C.A., Tepper R.I.;
RT "Identification and expression cloning of a leptin receptor, OB-R.";
RL Cell 83:1263-1271(1995).
[2]
SEQUENCE FROM N.A.
RP Thompson D.B., Ossowski V., Sutherland J., Apel W.,
RP Bielefeldt J.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
[3]
RP VARIANT ARG-223.
RX MEDLINE=96270489; PubMed=8666155;
RA Considine R.V., Considine E.L., Williams C.J., Hyde T.M., Caro J.F.;
RT "The hypothalamic leptin receptor in humans: identification of
RT incidental sequence polymorphisms and absence of the db/db mouse and
RT fa/fa rat mutations.";
RL Diabetes 45:992-994(1996).
[4]
RP VARIANTS ARG-109; ARG-204; ARG-223 AND ASN-656.
RX MEDLINE=97289527; PubMed=9144432;
RA Echwald S.M., Soerensen T.D., Soerensen T.I., Tybjaerg-Hansen A.,
RA Andersen T., Chung W.K., Leibel R.L., Pedersen O.;
RT "Amino acid variants in the human leptin receptor: lack of association
RT to juvenile onset obesity.";
RL Biochem. Biophys. Res. Commun. 233:248-252(1997).
[5]
RP VARIANTS ARG-109; ARG-223 AND ASN-656.
RX MEDLINE=97431549; PubMed=9287054;
RA Chung W.K., Power-Kehoe L., Chua M., Chu F., Aronne L., Huma 2.,
RA Sothorn M., Udall J.N., Kahle B., Leibel R.L.;
RT "Exonic and intronic sequence variation in the human leptin receptor
RT gene (LEPR).";
RL Diabetes 46:1509-1511(1997).
[6]
RP VARIANTS ARG-109 AND ARG-223.
RX MEDLINE=97301763; PubMed=9158141;
RA Thompson D.B., Ravussin E., Bennett P.H., Bogardus C.;
RT "Structure and sequence variation at the human leptin receptor gene in
RT lean and obese Pima Indians.";
RL Hum. Mol. Genet. 6:675-679(1997).
[7]
RP VARIANTS ARG-109; ARG-223; ASN-656 AND THR-675.
RX MEDLINE=99075638; PubMed=9860295;
RA Roth H., Korn T., Rosenkranz K., Hinney A., Ziegler A., Kunz J.,
RA Siegfried W., Mayer H., Hebebrand J., Grzeschik K.H.;
RT "Transmission disequilibrium and sequence variants at the leptin
RT receptor gene in extremely obese German children and adolescents.";
RL Hum. Genet. 103:540-546(1998).
[8]
RP VARIANTS ARG-109; ARG-223 AND ASN-656.
RX MEDLINE=97318795; PubMed=9175732;
RA Gotoda T., Manning B.S., Goldstone A.P., Imrie H., Evans A.L.,
RA Strosberg A.D., McKeigue P.M., Scott J., Altman T.J.;
RT "Leptin receptor gene variation and obesity: lack of association in a
RT white British male population.";

```

Hum. Mol. Genet. 6:869-876(1997).

CC -!- FUNCTION: RECEPTOR FOR OBESITY FACTOR (LEPTIN).

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- SIMILARITY: CONTAINS 3 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC -!- SIMILARITY: BELONGS TO THE CYTOKINE FAMILY OF RECEPTORS.

CC

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC

DR EMBL; U43168; AAA93015.1; -

DR EMBL; U59263; AAB09673.1; -

DR EMBL; U59248; AAB09673.1; JOINED.

DR EMBL; U59249; AAB09673.1; JOINED.

DR EMBL; U59250; AAB09673.1; JOINED.

DR EMBL; U59252; AAB09673.1; JOINED.

DR EMBL; U59253; AAB09673.1; JOINED.

DR EMBL; U59254; AAB09673.1; JOINED.

DR EMBL; U59255; AAB09673.1; JOINED.

DR EMBL; U59256; AAB09673.1; JOINED.

DR EMBL; U59257; AAB09673.1; JOINED.

DR EMBL; U59258; AAB09673.1; JOINED.

DR EMBL; U59259; AAB09673.1; JOINED.

DR EMBL; U59260; AAB09673.1; JOINED.

DR EMBL; U59261; AAB09673.1; JOINED.

DR EMBL; U59262; AAB09673.1; JOINED.

DR HSSP; P40189; 1BQU.

DR MIM; 601007; -

DR InterPro; IPR002996; CRIA.

DR InterPro; IPR003961; FN.III.

DR InterPro; IPR003529; Hematopo_receptor_L_F2.

DR Pfam; PF00041; fn3; 2.

DR SMART; SM00060; FN3; 1.

DR PROSITE; PS01353; HEMATOPO_REC_L_F2; 1.

KW Obesity; Receptor; Transmembrane; Glycoprotein; Signal; Repeat;

KW Polymorphism.

FT SIGNAL 1 21 POTENTIAL.

FT CHAIN 22 1165 LEPTIN RECEPTOR.

FT DOMAIN 22 841 EXTRACELLULAR (POTENTIAL).

FT TRANSMEM 842 862 POTENTIAL.

FT DOMAIN 863 1165 CYTOPLASMIC (POTENTIAL).

FT DOMAIN 237 320 FIBRONECTIN TYPE-III 1.

FT DOMAIN 537 623 FIBRONECTIN TYPE-III 2.

FT DOMAIN 738 823 FIBRONECTIN TYPE-III 3.

FT CARBOHYD 23 23 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 41 41 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 56 56 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 73 73 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 81 81 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 98 98 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 187 187 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 206 206 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 276 276 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 347 347 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 397 397 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 433 433 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 516 516 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 659 659 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 670 670 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 688 688 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 697 697 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 728 728 N-LINKED (GLCNAC. .) (POTENTIAL).

FT CARBOHYD 750 750 N-LINKED (GLCNAC. .) (POTENTIAL).

FT VARIANT 109 109 K -> R. /FTID-VAR_002703.

FT VARIANT 204 204 K -> R. /FTID-VAR_002704.

FT VARIANT 223 223 Q -> R.


```
DR EMBL; AE000529; AAD07147.1; -.
DR TIGR; HP0076; -.
DR InterPro; IPR002583; Ribosomal_S20p.
DR Pfam; PF01649; Ribosomal_S20p; 1.
DR ProDom; PD004231; Ribosomal_S20p; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 89 AA; 10196 MW; FD2DA520B39FEA36 CRC64;

Query Match 67.5%; Score 27; DB 1; Length 89;
Best Local Similarity 55.6%; Pred. No. 38;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
Db 78 RLNASVKKI 86
||||:||||:

RESULT 35
RLA3_CLAHE STANDARD; PRT; 111 AA.
AC P42038;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 60S acidic ribosomal protein P2 (Allergen Cla h 3) (Cla h III).
GN CLAH3.
OS Cladosporium herbarum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetothyriomycetes incertae sedis;
OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
OX NCBI_TaxID=29918;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-95114396; PubMed-7814877;
RA Zhang L., Muradta G., Curran I.H., Rode H., Vijay H.M.;
RT "A cDNA clone coding for a novel allergen, Cla h III, of Cladosporium
herbarum identified as a ribosomal P2 protein.";
RL J. Immunol. 154:710-717(1995).
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
PROTEIN SYNTHESIS.
CC -!- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
SUBUNIT.
CC -!- PTM: PHOSPHORYLATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; X77253; CAA54470.1; -.
DR InterPro; IPR001813; 60S_ribosomal.
DR Pfam; PF00428; 60S_ribosomal; 1.
KW Ribosomal protein; Phosphorylation; Multigene family; Allergen.
SQ SEQUENCE 111 AA; 11124 MW; 297FALL3FEDEF89 CRC64;

Query Match 67.5%; Score 27; DB 1; Length 111;
Best Local Similarity 55.6%; Pred. No. 46;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
Db 38 RLSSLKEL 46
||||:||||:

RESULT 36
RLA4_CLAHE STANDARD; PRT; 112 AA.
AC P42039;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 60S acidic ribosomal protein P2 (Allergen Cla h 4) (Cla h IV).
GN CLAH4.
OS Cladosporium herbarum.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina;
OC Dothideomycetes et Chaetothyriomycetes incertae sedis;
OC Mycosphaerellaceae; mitosporic Mycosphaerellaceae; Cladosporium.
OX NCBI_TaxID=29918;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=280202-BERLIN;
RX MEDLINE-95206305; PubMed=7898496;
RA Achatz G., Oberkofler H., Lechenauer E., Simon B., Unger A.,
RA Kandler D., Ebner C., Prillinger H., Kraft D., Breitenbach M.;
RT "Molecular cloning of major and minor allergens of Alternaria
alternata and Cladosporium herbarum.";
RL Mol. Immunol. 32:213-227(1995).
RN [2]
RP REVISIONS TO 13; 38-41 AND 93.
RA Simon-Nobbe B.;
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PLAYS AN IMPORTANT ROLE IN THE ELONGATION STEP OF
PROTEIN SYNTHESIS.
CC -!- SUBUNIT: P1 AND P2 EXIST AS DIMERS AT THE LARGE RIBOSOMAL
SUBUNIT.
CC -!- PTM: PHOSPHORYLATED (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE L12P FAMILY OF RIBOSOMAL PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; X78223; CAA55067.2; ALT-SEQ.
DR InterPro; IPR001813; 60S_ribosomal.
DR Pfam; PF00428; 60S_ribosomal; 1.
KW Ribosomal protein; Phosphorylation; Multigene family; Allergen.
SQ SEQUENCE 112 AA; 11234 MW; 2DC7B65C6AD997B7 CRC64;

Query Match 67.5%; Score 27; DB 1; Length 112;
Best Local Similarity 55.6%; Pred. No. 47;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
Db 38 RLSSLKEL 46
||||:||||:

RESULT 37
YB90_HAEIN STANDARD; PRT; 141 AA.
AC P44123;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein H1190.
GN H1190.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE-95350630; PubMed-7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
```

```

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus
RA Influenzae Rd.";
RA Science 269:496-512(1995).
CC -!- SIMILARITY: BELONGS TO THE PTPS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32798; AAC22843.1; -
DR TIGR; H11190; -
DR InterPro; IPR001914; PTPS.
DR Pfam; PF01242; PTPS; 1.
DR ProDom; PD004049; PTPS; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 141 AA; 16322 MW; 6843517DEA5A7C3E CRC64;
Query Match 67.5%; Score 27; DB 1; Length 141;
Best Local Similarity 75.0%; Pred. No. 58;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LSSMKVKV 9
Db 57 LKSIKVKV 64
| | | | |
| | | | |
RESULT 38
HSCB_ECOLI
ID HSCB_ECOLI STANDARD; PRT; 171 AA.
AC P36540;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chaperone protein hscB (HSC20).
GN HSCB OR B2527 OR Z3794 OR ECS3393.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_taxid=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=94131939; PubMed=8300516;
RA Kawula T.H., Lelievre M.J.;
RT "Mutations in a gene encoding a new Hsp70 suppress rapid DNA
RT inversion and bgl activation, but not proU derepression, in hns-1
RT mutant Escherichia coli.";
RL J. Bacteriol. 176:610-619(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregory J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,

```

```

RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobo T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kunara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP SEQUENCE OF 119-171 FROM N.A.
RC STRAIN=B;
RX MEDLINE=94181532; PubMed=8134349;
RA Seaton B.L., Vickery L.E.;
RT "A gene encoding a DnaK/hsp70 homolog in Escherichia coli.";
RL Proc. Natl. Acad. Sci. U.S.A. 91:2066-2070(1994).
RN [7]
RP CHARACTERIZATION.
RX MEDLINE=97289991; PubMed=9144776;
RA Vickery L.E., Silberg J.J., Ta D.T.;
RT "Hsc66 and Hsc20, a new heat shock cognate molecular chaperone system
RT from Escherichia coli.";
RL Protein Sci. 6:1047-1056(1997).
RN [8]
RP CRYSTALLIZATION.
RX MEDLINE=97445617; PubMed=9300502;
RA Cupp-Vickery J.R., Vickery L.E.;
RT "Crystallization and preliminary x-ray crystallographic properties of
RT Hsc20, a J-motif co-chaperone protein from Escherichia coli.";
RL Protein Sci. 6:2028-2030(1997).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS).
RX MEDLINE=20574017; PubMed=11124030;
RA Cupp-Vickery J.R., Vickery L.E.;
RT "Crystal structure of Hsc20, a J-type co-chaperone from Escherichia
RT coli.";
RL J. Mol. Biol. 304:835-845(2000).
CC -!- FUNCTION: CO-CHAPERONE THAT INTERACTS WITH HSCA AND STIMULATES
CC ITS ATPASE ACTIVITY.
CC -!- SIMILARITY: CONTAINS 1 J DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U01827; AAA18299.1; -

```

```
DR EMBL; AE000339; AAC75580.1; -.
DR EMBL; D90883; BAA16421.1; -.
DR EMBL; AE005482; AAG57641.1; -.
DR EMBL; AP002562; BAB36816.1; -.
DR EMBL; U05338; -. NOT_ANNOTATED_CDS.
DR PDB; 1FPO; 08-DEC-00.
DR EcoGene; EG12131; hscB.
DR InterPro; IPR001623; DnaJ_N.
DR Pfam; PF00226; DnaJ; 1.
DR SMART; SM00271; DnaJ; 1.
DR PROSITE; PS00636; DnaJ_1; FALSE_NEG.
DR PROSITE; PS00076; DnaJ_2; 1.
KW Chaperone; 3D-structure; Complete proteome.
FT DOMAIN 2 60 J-DOMAIN.
SQ SEQUENCE 171 AA; 20138 MW; 095193EE98AA60C9 CRC64;

Query Match 67.5%; Score 27; DB 1; Length 171;
Best Local Similarity 55.6%; Pred. No. 70;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
DB 113 RLESFIKRV 121

RESULT 39
ID PSAP_GUITH STANDARD; PRT; 183 AA.
AC 078457;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Photosystem I reaction centre subunit III (PSI-F).
PSAP.
GN Guillardia theta (Cryptomonas phi).
OS Guillardia theta (Cryptomonas phi).
OC Chloroplast.
OG Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=99128221; PubMed=9929392;
RA Douglas S.E., Penny S.L.;
RT "The plastid genome of the cryptophyte alga, Guillardia theta: complete sequence and conserved syntenic groups confirm its common ancestry with red algae.";
RT J. Mol. Evol. 48:236-244(1999).
CC -!- FUNCTION: PROBABLY PARTICIPATES IN EFFICIENCY OF ELECTRON TRANSFER FROM PLASTOCYANIN TO P700 (OR CYTOCHROME C553 IN ALGAE AND CYANOBACTERIA). THIS PLASTOCYANIN-DOCKING PROTEIN CONTRIBUTES TO THE SPECIFIC ASSOCIATION OF PLASTOCYANIN TO PS I.
CC -!- SUBCELLULAR LOCATION: ASSOCIATED WITH LUMINAL SIDE OF THE THYLAKOID MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE PSAP FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
DR EMBL; AF041468; AAC35648.1; -.
DR InterPro; IPR003666; PSI_Psaf.
DR Pfam; PF02507; PSI_Psaf; 1.
KW Photosynthesis; Photosystem I; Chloroplast; Thylakoid; Membrane.
SQ SEQUENCE 183 AA; 20517 MW; 924E5DBFAE58C9EF CRC64;

Query Match 67.5%; Score 27; DB 1; Length 183;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
```

```
DB 39 RLSSSVKKL 47

RESULT 40
Y101_MYCGE
ID Y101_MYCGE STANDARD; PRT; 222 AA.
AC P47347; O49244;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG101.
GN MG101.
OS Mycoplasma genitalium.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2097;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=96026346; PubMed=7569993;
RA Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A., Frieschmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M., Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L., Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M., Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S., Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT "The minimal gene complement of Mycoplasma genitalium.";
RL Science 270:397-403(1995).
[2]
RN SEQUENCE OF 27-132 FROM N.A.
RC STRAIN=ATCC 33530 / G-37;
RX MEDLINE=94075230; PubMed=8253680;
RA Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT "A survey of the Mycoplasma genitalium genome by using random sequencing.";
RL J. Bacteriol. 175:7918-7930(1993).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
DR EMBL; U39690; AAC71319.1; -.
DR EMBL; U02103; AAD12375.1; -.
DR TIGR; MG101; -.
KW Hypothetical protein; Complete proteome.
FT CONFLICT 27 48 WPANTRIFSERQLEIRFNSSRS -> HPYLLRTPTRNPFQL LTD (IN REF. 2).
FT SEQUENCE 222 AA; 25998 MW; 9D06A5480E58124D CRC64;

Query Match 67.5%; Score 27; DB 1; Length 222;
Best Local Similarity 44.4%; Pred. No. 90;
Matches 4; Conservative 5; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
DB 96 KLSTLIKKL 104

RESULT 41
YCX3_CHLVU
ID YCX3_CHLVU STANDARD; PRT; 250 AA.
AC O20130;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 28.3 kDa protein in rrn23-psbC intergenic region (ORF250).
OS Chlorella vulgaris.
```

OG Chloroplast.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Chlorella.
OX NCBI_TaxID=3077;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IAM C-27 / TAMIYA;
RX MEDLINE=97303241; PubMed=9159184;
RA Wakasugi T., Nagai T., Kapoor M., Sugita M., Ito S.,
RA Tsudzuki J., Nakashima K., Tsudzuki T., Suzuki Y., Hamada A., Ohta T.,
RA Inamura A., Yoshinaga K., Suglura M.
RT "Complete nucleotide sequence of the chloroplast genome from the green
RT alga Chlorella vulgaris: the existence of genes possibly involved in
RT chloroplast division."
RL Proc. Natl. Acad. Sci. U.S.A. 94:5967-5972(1997).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB001684; BAA57870.1; -
KW Hypothetical protein; Chloroplast.
SQ SEQUENCE 250 AA; 28261 MW; 4A23DF01790A9677 CRC64;

Query Match 67.5%; Score 27; DB 1; Length 250;
Best Local Similarity 62.5%; Pred. No. 1.le+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RLSSMVKK 8
Db 131 RLSNLLKK 138
||||:|

RESULT 42
PROB_STRPTR STANDARD; PRT; 267 AA.
AC P96488;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Glutamate 5-kinase (EC 2.7.2.11) (Gamma-glutamyl kinase) (GK).
GN PROB.
OS Streptococcus thermophilus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1308;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19258;
RX MEDLINE=97124211; PubMed=8969524;
RA Limauro D., Falcitatore A., Basso A.L., Forlani G., de Felice M.;
RT "Proline biosynthesis in Streptococcus thermophilus: characterization
RT of the proBA operon and its products."
RL Microbiology 142:3275-3282(1996).
CC -!- FUNCTION: Catalyzes the transfer of a phosphate group to glutamate
CC to form glutamate 5-phosphate which rapidly cyclizes to 5-
CC oxoproline.
CC -!- CATALYTIC ACTIVITY: ATP + L-glutamate = ADP + L-glutamate 5-
CC phosphate.
CC -!- PATHWAY: proline biosynthesis pathway; first step.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (by similarity).
CC -!- SIMILARITY: BELONGS TO THE GLUTAMATE 5-KINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

CC or send an email to license@isb-sib.ch.
CC -----
DR EMBL; X92418; CAA63147.1; -
DR InterPro; IPR001048; Aakkinase.
DR InterPro; IPR001057; Glut_5_kinase.
DR Pfam; PF00696; aakkinase; 1.
DR PRINTS; PR00474; GLU5KINASE.
DR PROSITE; PS00902; GLUTAMATE_5_KINASE; 1.
KW Transferase; Kinase; Proline biosynthesis.
SQ SEQUENCE 267 AA; 28966 MW; B72448C2B7432EE3 CRC64;

Query Match 67.5%; Score 27; DB 1; Length 267;
Best Local Similarity 66.7%; Pred. No. 1.le+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLSSMVKKV 9
Db 159 RLSAIVAKV 167
||||:|

RESULT 43
RCN2_HUMAN STANDARD; PRT; 317 AA.
ID RCN2_HUMAN
AC Q14257;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Reticulocalbin 2 precursor (Calcium-binding protein ERC-55) (E6-
DE binding protein) (EBBP).
GN RCN2 OR ERC55.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94308182; PubMed=8034671;
RA Weis K., Griffiths G., Lamond A.I.;
RT "The endoplasmic reticulum calcium-binding protein of 55 kDa is a
RT novel EF-hand protein retained in the endoplasmic reticulum by a
RT carboxyl-terminal His-Asp-Glu-Leu motif."
RL J. Biol. Chem. 269:19142-19150(1994).
CC -!- FUNCTION: NOT KNOWN, BINDS CALCIUM.
CC -!- SUBCELLULAR LOCATION: Endoplasmic reticulum lumen.
CC -!- TISSUE SPECIFICITY: UBIQUITOUS.
CC -!- SIMILARITY: BELONGS TO THE RETICULOCALBIN FAMILY.
CC -!- SIMILARITY: CONTAINS 6 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X78669; CAA55343.1; -
DR HSP; P02633; IBOC.
DR MIM; 602584; -
DR InterPro; IPR002048; EF-hand.
DR InterPro; IPR000886; EF_target.
DR Pfam; PF00036; ehand; 5.
DR SMART; SM00054; EFh; 2.
DR PROSITE; PS00014; EF_TARGET; 1.
DR PROSITE; PS00018; EF_HAND; 5.
KW Calcium-binding; Endoplasmic reticulum; Signal; Repeat.
FT SIGNAL 1 22
FT CHAIN 23 317
FT RETICULOCALBIN 2.
FT CA_BIND 74 85
FT EF-HAND 1 (POTENTIAL).
FT CA_BIND 110 121
FT EF-HAND 2 (POTENTIAL).
FT CA_BIND 162 173
FT EF-HAND 3 (POSSIBLY ANCESTRAL).
FT CA_BIND 199 210
FT EF-HAND 4 (POTENTIAL).
FT CA_BIND 240 251
FT EF-HAND 5 (POTENTIAL).

```
FT CA_BIND 276 287 EF-HAND 6 (POTENTIAL).
SQ SITE 314 317 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 317 AA: 36876 MW; E7C0A4211DBF04AB CRC64;

Query Match 67.5%; Score 27; DB 1; Length 317;
Best Local Similarity 44.4%; Pred. No. 1.3e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMKVK 9
DB 65 RLQAIKKI 73

RESULT 44
TRXB_BUCAI STANDARD; PRT; 319 AA.
AC P57399;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Thioredoxin reductase (EC 1.6.4.5) (TRXR).
GN TRXB OR BU314.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. APS."
RL Nature 407:81-86(2000).
CC -|- CATALYTIC ACTIVITY: NADPH + oxidized thioredoxin = NADP(+) + reduced thioredoxin.
CC -|- COPACITOR: FAD (BY SIMILARITY).
CC -|- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC -|- MISCELLANEOUS: THE ACTIVE SITE IS A REDOX-ACTIVE DISULFIDE BOND.
CC -|- SIMILARITY: BELONGS TO THE PYRIDINE NUCLEOTIDE-DISULFIDE OXIDOREDUCTASES CLASS-II.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; AP001119; BAB13022.1;
DR InterPro; IPR000759; Adnrx_reductase.
DR InterPro; IPR001327; FAD_pyr_redox.
DR InterPro; IPR000103; Pyridine_redox_2.
DR InterPro; IPR001100; pyr_redox.
DR Pfam; PF00070; pyr_redox; 1.
DR PRINTS; PR00419; ADXRDTASE.
DR PRINTS; PR00368; FADPNR.
DR PRINTS; PR00411; PNDRTASEI.
DR PRINTS; PR00469; PNDRTASEII.
DR PROSITE; PS00573; PYRIDINE_REDOX_2; 1.
KW Redox-active center; Oxidoreductase; NADP; Flavoprotein; FAD;
Complete proteome.
FT NP_BIND 7 22 FAD (ADP PART) (PROBABLE).
FT DISULFID 136 139 REDOX-ACTIVE.
FT NP_BIND 148 162 NAD(P) (BY SIMILARITY).
FT NP_BIND 278 288 FAD (FLAVIN PART) (BY SIMILARITY).
SQ SEQUENCE 319 AA; 35105 MW; 27095618CCF1E0F6 CRC64;

Query Match 67.5%; Score 27; DB 1; Length 319;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

FT CA_BIND 276 287 EF-HAND 6 (POTENTIAL).
SQ SITE 314 317 PREVENT SECRETION FROM ER (POTENTIAL).
SQ SEQUENCE 317 AA: 36876 MW; E7C0A4211DBF04AB CRC64;

Query Match 67.5%; Score 27; DB 1; Length 317;
Best Local Similarity 44.4%; Pred. No. 1.3e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMKVK 9
DB 65 RLQAIKKI 73

RESULT 45
XYLB_PSEPU STANDARD; PRT; 366 AA.
AC P39849;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Aryl-alcohol dehydrogenase (EC 1.1.1.90) (Benzyl alcohol dehydrogenase) (BADH).
GN XYLB.
OS Pseudomonas putida.
OG Plasmid TOL pMW0, and Plasmid TOL pMW53.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=303;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-8.
RC PLASMID=TOL pMW0;
RX MEDLINE=93266524; PubMed=8496150;
RA Shaw J.P., Rekik M., Schwager F., Harayama S.;
RT "Kinetic studies on benzyl alcohol dehydrogenase encoded by TOL plasmid pMW0. A member of the zinc-containing long chain alcohol dehydrogenase family."
RL J. Biol. Chem. 268:10842-10850(1993).
RN [2]
RP SEQUENCE OF 1-53.
RC PLASMID=TOL pMW53;
RX MEDLINE=91113163; PubMed=1989592;
RA Chalmers R.M., Keen J.N., Fewson C.A.;
RT "Comparison of benzyl alcohol dehydrogenases and benzaldehyde dehydrogenases from the benzyl alcohol and mandelate pathways in Acinetobacter calcoaceticus and from the TOL-plasmid-encoded toluene pathway in Pseudomonas putida. N-terminal amino acid sequences, amino acid compositions and immunological cross-reactions."
RL Biochem. J. 273:99-107(1991).
CC -|- FUNCTION: OXIDIZES PRIMARY ALCOHOLS WITH AN AROMATIC OR CYCLOHEX-1-ENE RING. IT IS HIGHLY SPECIFIC FOR BENZYL ALCOHOL.
CC -|- CATALYTIC ACTIVITY: An aromatic alcohol + NAD(+) = an aromatic aldehyde + NADH.
CC -|- COPACITOR: ZINC; BINDS 2 ZINC IONS PER SUBUNIT.
CC -|- SUBUNIT: HOMODIMER.
CC -|- SIMILARITY: BELONGS TO THE ZINC-CONTAINING ALCOHOL DEHYDROGENASE FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; D63341; BAA09664.1;
DR EMBL; M94184; AAA26024.1; ALT_INIT.
DR PIR; A46704; A46704.
DR HSSP; P00325; 1DEH.
DR InterPro; IPR002328; ADH_zinc.
DR InterPro; IPR002085; Adh_zn_family.
DR InterPro; IPR000205; NAD_binding.
DR Pfam; PF00107; adh_zinc; 1.
DR PROSITE; PS00059; ADH_ZINC; 1.
KW Aromatic hydrocarbons catabolism; Oxidoreductase; NAD; Plasmid.
FT METAL 40 40 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 61 61 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 90 90 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 93 93 ZINC (SECOND ATOM) (BY SIMILARITY).
FT METAL 96 96 ZINC (SECOND ATOM) (BY SIMILARITY).
```

FT METAL 104 104 ZINC (SECOND ATOM) (BY SIMILARITY).
 FT METAL 169 169 ZINC (CATALYTIC) (BY SIMILARITY).
 FT CONFLICT 40 41 CH -> PG (IN REF. 2).
 FT CONFLICT 50 50 H -> G (IN REF. 2).
 SQ SEQUENCE 366 AA; 38510 MW; 23A5203C0F96D91B CRC64;

Query Match 67.5%; Score 27; DB 1; Length 366;
 Best Local Similarity 66.7%; Pred. NO. 1.5e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSWVKV 9
 Db 69 RVGSAVKV 77

Search completed: November 6, 2002, 12:09:04
 Job time : 8.3333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 5.33333 Seconds
(without alignments)
65.339 Million cell updates/sec

Title: US-09-905-083-33
Perfect score: 40
Sequence: 1 LLLPLQILL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	40	100.0	253	1 KLK7_HUMAN	P49862 homo sapien
2	38	95.0	361	1 IHA_TRITU	O77755 trichosurus
3	34	85.0	369	1 PSPD_BOVIN	P35246 bos taurus
4	34	85.0	371	1 CONG_BOVIN	P23805 bos taurus
5	33	82.5	428	1 YFJD_ECOLI	P37908 escherichia
6	33	82.5	435	1 YA06_HUMAN	O60813 homo sapien
7	33	82.5	500	1 YA04_HUMAN	O60810 homo sapien
8	33	82.5	653	1 APP1_MOUSE	Q03157 mus musculus
9	32	80.0	198	1 CD8A_PONPY	P30433 pongo pygma
10	32	80.0	235	1 CD8A_HUMAN	P10732 homo sapien
11	32	80.0	238	1 MABA_RAT	P19999 rattus norv
12	32	80.0	239	1 CD8A_CANFA	P33706 canis famli
13	32	80.0	247	1 MCT1_HUMAN	P23946 homo sapien
14	32	80.0	247	1 MCT1_MACFA	P56435 macaca fasc
15	32	80.0	247	1 MCT1_PAPHA	P52195 papio hamad
16	32	80.0	492	1 MM11_MOUSE	Q02853 mus musculus
17	32	80.0	1306	1 ACE_HUMAN	P12821 homo sapien
18	31	77.5	27	1 ACH4_MOUSE	O70174 mus musculus
19	31	77.5	100	1 APC2_CAVPO	P27916 cavia porce
20	31	77.5	153	1 IL2_RABIT	O77620 oryctolagus
21	31	77.5	155	1 IL2_RAT	P17108 rattus norv
22	31	77.5	184	1 MUPM_MOUSE	P04939 mus musculus
23	31	77.5	249	1 YGRS_STRCO	P40179 streptomyce
24	31	77.5	338	1 YD4_SCHPO	P87052 schizosacch
25	31	77.5	461	1 NCBI_HUMAN	Q02818 homo sapien
26	31	77.5	499	1 C722_ARATH	Q91xm3 arabidopsis
27	31	77.5	592	1 NDC1_HUMAN	Q13183 homo sapien
28	31	77.5	625	1 TR11_MOUSE	Q35305 mus musculus
29	31	77.5	630	1 ACH4_RAT	P09483 rattus norv
30	31	77.5	650	1 APPE_HUMAN	P51693 homo sapien
31	31	77.5	699	1 PTPE_MOUSE	P49446 mus musculus
32	31	77.5	700	1 PTPE_HUMAN	P23469 homo sapien
33	31	77.5	1007	1 CHC2_HUMAN	Q9Y514 homo sapien

RESULT 1				
KLK7_HUMAN	STANDARD;	PRT;	253 AA.	
ID	KLK7_HUMAN			
AC	P49862;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic enzyme) (hSCCE).			
DE	KLK7 OR PRSS6 OR SCCE.			
GN	KLK7 OR PRSS6 OR SCCE.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.			
RC	TISSUE=Skin;			
RX	MEDLINE=94308225; PubMed=8034709;			
RA	Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A.,			
RA	Egelund T.;			
RT	"Cloning, expression, and characterization of stratum corneum			
RT	chymotryptic enzyme. A skin-specific human serine proteinase.";			
RL	J. Biol. Chem. 269:19420-19426(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RP	TISSUE=Keratinocytes;			
RA	Yousef G.M., Scorilas A., Diamandis E.P.;			
RT	"Molecular characterization, mapping and tissue expression of the			
RT	human stratum corneum chymotryptic enzyme gene.";			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20510030; PubMed=11054574;			
RA	Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,			
RA	Moss P., Paepier B., Wang K.;			
RT	"Sequencing and expression analysis of the serine protease gene			
RT	cluster located in chromosome 19q13 region.";			
RL	Gene 257:119-130(2000).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=95314630; PubMed=7794273;			
RA	Skytt A., Stroemqvist M., Egelund T.;			
RT	"Primary substrate specificity of recombinant human stratum corneum			
RT	chymotryptic enzyme.";			
RL	Biochem. Biophys. Res. Commun. 211:586-589(1995).			
CC	!- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE			
CC	STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS			
CC	SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID			
CC	RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE			
CC	CLEAVES INSULIN B CHAIN AT 6-LEU-I-CYS-7, 16-TYR-I-LEU-17, 25-PHE-			
CC	I-TYR-26, AND 26-TYR-I-THR-27. COULD PLAY A ROLE IN THE ACTIVATION			
CC	OF PRECURSORS TO INFLAMMATORY CYTOKINES.			
CC	!- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS			
CC	EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE			
CC	ALSO SEEN IN THE BRAIN AND KIDNEY.			
CC	!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			

p97798 mus musculus
Q961q0 homo sapien
P98161 homo sapien
P43030 sus scrofa
P26994 pseudomonas
P26891 sus scrofa
Q05016 bos taurus
Q29416 canis famli
P36835 capra hircu
P19114 ovis aries
P51747 cervus elap
Q00298 botrytis ci

```
CC TRYPsin FAMILY, KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L33404; AAC37551.1; -
DR EMBL: AF166330; AAD49718.1; -
DR EMBL: AF243527; AAG33360.1; -
DR HSSP: P00763; LDPO.
DR MEROPS: S01.300; -.
DR MIM: 604438; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal.
FT SIGNAL 1 22
FT PROPEP 23 29 ACTIVATION PEPTIDE.
FT CHAIN 30 253 KALLIKREIN 7.
FT ACT_SITE 70 70 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 36 137 BY SIMILARITY.
FT DISULFID 55 71 BY SIMILARITY.
FT DISULFID 137 239 BY SIMILARITY.
FT DISULFID 144 211 BY SIMILARITY.
FT DISULFID 176 190 BY SIMILARITY.
FT DISULFID 201 226 BY SIMILARITY.
FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 253 AA; 27525 MW; 2D68B6B15A76A668 CRC64;

Query Match 100.08; Score 40; DB 1; Length 253;
Best Local Similarity 100.08; Pred. No. 0.88;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
Db 5 LLLPLQLLL 13
|||||

RESULT 2
ID IHA_TRIVU STANDARD; PRT; 361 AA.
AC Q77755;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Inhibin alpha chain precursor.
GN INHA.
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
OX NCBI_TaxID=9337;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99027340; PubMed=9801457;
RA Vamontfort D., Fidler A.E., Heath D.A., Lawrence S.B., Tisdall D.J.,
RA Greenwood P.J., McNatty K.;
RT "cDNA sequence analysis, gene expression and protein localisation of
RT the inhibin alpha subunit of Australian brushtail possum (Trichosurus
RT vulpecula).";
RL J. Mol. Endocrinol. 21:141-152(1998).
CC -!- FUNCTION: INHIBINS AND ACTIVINS INHIBIT AND ACTIVATE,
CC RESPECTIVELY, THE SECRETION OF FOLLITROPIN BY THE PITUITARY GLAND.
```

```
CC INHIBINS/ACTIVINS ARE INVOLVED IN REGULATING A NUMBER OF DIVERSE
CC FUNCTIONS SUCH AS HYPOTHALAMIC AND PITUITARY HORMONE SECRETION,
CC GONADAL HORMONE SECRETION, GERM CELL DEVELOPMENT AND MATURATION,
CC ERYTHROID DIFFERENTIATION, INSULIN SECRETION, NERVE CELL SURVIVAL,
CC EMBRYONIC AXIAL DEVELOPMENT OR BONE GROWTH, DEPENDING ON THEIR
CC SUBUNIT COMPOSITION. INHIBINS APPEAR TO OPPOSE THE FUNCTIONS OF
CC ACTIVINS.
CC -!- SUBUNIT: DIMERIC, LINKED BY ONE OR MORE DISULFIDE BONDS.
CC INHIBIN A IS A DIMER OF ALPHA AND BETA-A.
CC INHIBIN B IS A DIMER OF ALPHA AND BETA-B.
CC -!- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF033340; AAC63945.1; -
DR InterPro: IPR002405; Inhibin_alpha.
DR InterPro: IPR001839; TGF-beta.
DR Pfam: PF00019; TGF-beta; 1.
DR PRINTS: PR00669; INHIBINA.
DR ProDom: PD000357; TGF-beta; 1.
DR SMART: SM00204; TGF-beta; 1.
DR PROSITE: PS00250; TGF_BETA_1; 1.
KW Growth factor; Hormone; Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 230 BY SIMILARITY.
FT CHAIN 231 361 INHIBIN ALPHA CHAIN.
FT DISULFID 260 323 BY SIMILARITY.
FT DISULFID 289 358 BY SIMILARITY.
FT DISULFID 293 360 BY SIMILARITY.
FT DISULFID 322 322 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 361 AA; 38945 MW; D661CDF93CDAA87D CRC64;

Query Match 95.0%; Score 38; DB 1; Length 361;
Best Local Similarity 88.9%; Pred. No. 3.1;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
Db 4 LLLPLQLLL 12
|||||

RESULT 3
ID PSPD_BOVIN STANDARD; PRT; 369 AA.
AC P35246;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN SFTPD OR SFTPD.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.
RC TISSUE=Lung;
RX MEDLINE=93170856; PubMed=8436402;
RA Lim B.L., Lu J., Reid K.B.M.;
RT "Structural similarity between bovine conglutinin and bovine lung
RT surfactant protein D and demonstration of liver as a site of
RT synthesis of conglutinin.";
RL Immunology 78:159-165(1993).
```


CC -!- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
 CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
 CC EXTENT OTHER ALPHA-GLUCOSYL MOIETIES. IT COULD PARTICIPATE IN THE
 CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -!- CELLULAR LOCATION: Extracellular.
 CC -!- MISCELLANEOUS: PULMONARY SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X75911; CA53510.1;
 CC PIR: S33603; S33603.
 CC DR HSP; P35247; I808.
 CC DR InterPro: IPR000087; Collagen.
 CC DR InterPro: IPR001304; lectin_c.
 CC DR Pfam: PF01391; Collagen; 3.
 CC DR Pfam: PF00059; lectin_c; 1.
 CC DR SMART; SM00034; CLECT; 1.
 CC DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
 CC DR PROSITE; PS50041; C-TYPE LECTIN_2; 1.
 CC KW Glycoprotein; Calcium; Surface film; Gaseous exchange; Hydroxylation;
 KW Signal; Lectin; Collagen; Repeat; Coiled coil.
 FT SIGNAL 1 20
 FT CHAIN 21 369
 FT
 FT DOMAIN 46 216
 FT COLLAGEN-LIKE.
 FT COILED COIL (POTENTIAL).
 FT DOMAIN 217 248
 FT C-TYPE LECTIN (SHORT FORM).
 FT DOMAIN 273 369
 FT DISULFID 275 367
 FT BY SIMILARITY.
 FT DISULFID 345 359
 FT BY SIMILARITY.
 FT CARBOHYD 90 90
 FT N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT MOD_RES 78 78
 FT HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 87 87
 FT HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 96 96
 FT HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 99 99
 FT HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 165 165
 FT HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 171 171
 FT HYDROXYLATION (BY SIMILARITY).
 SQ SEQUENCE 369 AA; 37361 MW; 07D88B24E0AEB2E3 CRC64;
 Query Match 85.0%; Score 34; DB 1; Length 369;
 Best Local Similarity 77.8%; Pred. No. 19;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPLQLL 9
 Db 2 LLLPLSVLL 10
 ||||| :||
 ||||| :||
 RESULT 4
 CONG_BOVIN STANDARD; PRT; 371 AA.
 ID CONG_BOVIN STANDARD; PRT; 371 AA.
 AC P23805;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Conglutinin precursor.
 GN CGN1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
 OC Bovidae; Bovinae; Bos.
 OC NCBI_TaxID=9913;

RN SEQUENCE FROM N.A.
 RP MEDLINE=93213261; PubMed=8460993;
 RX Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
 RA "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
 RT Biochem. Biophys. Res. Commun. 191:335-342(1993).
 RL
 RN SEQUENCE FROM N.A.
 RP MEDLINE=93277452; PubMed=7684896;
 RX Lu J., Laursen S.B., Thiel S., Jensenius J.C., Reid B.M.;
 RA "The cDNA cloning of conglutinin and identification of liver as a
 RT primary site of synthesis of conglutinin in members of the Bovidae.";
 RL Biochem. J. 292:157-162(1993).
 RN SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94215917; PubMed=8163202;
 RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,
 RT Tauber A.I., Sastry K.N.;
 "Bovine conglutinin (BC) mRNA expressed in liver: cloning and
 RT characterization of the BC cDNA reveals strong homology to surfactant
 RT protein-D.";
 RL Gene 141:277-281(1994).
 RN SEQUENCE FROM N.A.
 RP SEQUENCE FROM N.A.
 RC TISSUE=Semen;
 RX MEDLINE=94267222; PubMed=8207234;
 RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,
 RT Tauber A.I., Sastry K.N.;
 "Bovine conglutinin gene exon structure reveals its evolutionary
 RT relationship to surfactant protein-D.";
 RL J. Immunol. 153:173-180(1994).
 RN SEQUENCE OF 21-371.
 RP MEDLINE=91131556; PubMed=1993651;
 RX Lee Y.-M., Leiby K.R., Allar J., Paris K., Lerch B., Okarma T.B.;
 RA "Primary structure of bovine conglutinin, a member of the C-type
 RT animal lectin family.";
 RL J. Biol. Chem. 266:2715-2723(1991).
 RN PRELIMINARY SEQUENCE OF 21-52.
 RP MEDLINE=87184551; PubMed=3556740;
 RX Young N.M., Leon M.A.;
 RA "The carbohydrate specificity of conglutinin and its homology to
 RT proteins in the hepatic lectin family.";
 RL Biochem. Biophys. Res. Commun. 143:645-651(1987).
 CC -!- FUNCTION: CALCIUM-DEPENDENT LECTIN-LIKE PROTEIN WHICH BINDS TO A
 CC YEAST CELL WALL EXTRACT & IMMUNE COMPLEXES THROUGH THE COMPLEMENT
 CC COMPONENT (C3BI). IT IS CAPABLE OF BINDING NONREDUCING TERMINAL
 CC N-ACETYLGLUCOSAMINE, MANNOSE, AND FUCOSE RESIDUES.
 CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: D14085; BAA03170.1;
 CC EMBL: X71774; CAA50665.1;
 CC EMBL: L18871; AAA20126.1;
 CC EMBL: U06860; AAB60624.1;
 CC EMBL: U06854; AAB60624.1; JOINED.
 CC EMBL: U06855; AAB60624.1; JOINED.
 CC EMBL: U06856; AAB60624.1; JOINED.
 CC EMBL: U06857; AAB60624.1; JOINED.
 CC EMBL: U06858; AAB60624.1; JOINED.
 CC EMBL: U06859; AAB60624.1; JOINED.
 CC PIR: A29416; A29416.

```

DR PIR; A23740; A23740.
DR HSP; JN0450; JN0450.
DR HSP; P35247; 1B08.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE; PS00041; C-TYPE-LECTIN_2; 1.
KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;
KW Collagen; Repeat; Calcium; Signal.
FT SIGNAL 1
FT CHAIN 21 371 CONGLUTININ.
FT DOMAIN 46 216 COLLAGEN-LIKE.
FT DOMAIN 273 371 C-TYPE LECTIN (SHORT FORM).
FT MOD_RES 63 63 HYDROXYLATION.
FT MOD_RES 87 87 HYDROXYLATION.
FT MOD_RES 99 99 HYDROXYLATION.
FT MOD_RES 135 135 HYDROXYLATION.
FT MOD_RES 141 141 HYDROXYLATION.
FT MOD_RES 159 159 HYDROXYLATION.
FT MOD_RES 162 162 HYDROXYLATION.
FT MOD_RES 198 198 HYDROXYLATION.
FT SITE 201 203 CELL ATTACHMENT SITE (POTENTIAL).
FT DISULFID 275 369 BY SIMILARITY.
FT CARBOHYD 337 337 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 173 173 R -> H (IN REF. 2 AND 3).
FT CONFLICT 210 210 K -> S (IN REF. 5).
FT CONFLICT 218 218 V -> A (IN REF. 2).
FT CONFLICT 272 272 E -> V (IN REF. 2).
SQ SEQUENCE 371 AA; 37994 MW; 867BBA1992544B1F CRC64;

Query Match 85.0%; Score 34; DB 1; Length 371;
Best Local Similarity 77.8%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 2 LLLPLSVLL 10

RESULT 5
ID YFJD_ECOLI STANDARD; PRT; 428 AA.
AC P37908; P76600; P76601; P77009;
DT 01-OCT-1994 (Rel. 30; Created)
DT 15-DEC-1998 (Rel. 37; Last sequence update)
DE Hypothetical protein yfjd.
GN YFJD OR B2612/B2613.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yanamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,

RA PIR; A23740; A23740.
RA PIR; JN0450; JN0450.
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-89.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [3]
RP SEQUENCE OF 322-428 FROM N.A.
RC STRAIN-B178;
RX MEDLINE=88319942; PubMed=3045760;
RA Lipinska B., King J., Ang D., Georgopoulos C.;
RT "Sequence analysis and transcriptional regulation of the Escherichia
RT coli grpE gene, encoding a heat shock protein.";
RL Nucleic Acids Res. 16:7545-7562(1988).
RN [4]
RP IDENTIFICATION.
RA Rued K.E.;
RL Unpublished observations (AUG-1994).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -!- SIMILARITY: BELONGS TO THE UPF0053 FAMILY. STRONG, TO
CC H-INFLUENZA H10107.
CC -!- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 199 THAT PRODUCES TWO ORFS.
CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO NUMEROUS
CC FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AE000347; AAC75661.1; ALT_FRAME.
DR EMBL; AE000347; AAC75662.1; ALT_FRAME.
DR EMBL; D90888; BAA16497.1; ALT_INIT.
DR EMBL; X07863; -; NOT_ANNOTATED_CDS.
DR EcoGene; EG12442; Yfjd.
DR InterPro; IPR000644; CBS.
DR Pfam; PF00571; CBS; 2.
DR Pfam; PF01595; DUF21; 1.
DR SMART; SM00116; CBS; 1.
KW Hypothetical protein; CBS domain; Repeat; Transmembrane;
KW Complete proteome.
FT TRANSMEM 4 24 POTENTIAL.
FT TRANSMEM 65 85 POTENTIAL.
FT TRANSMEM 92 112 POTENTIAL.
FT TRANSMEM 130 150 POTENTIAL.
SQ SEQUENCE 428 AA; 48044 MW; 38456865EDBCB151 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 428;
Best Local Similarity 77.8%; Pred. No. 34;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 131 LLAPLQILL 139

RESULT 6
ID YAO6_HUMAN STANDARD; PRT; 435 AA.
AC O60813;
DT 30-MAY-2000 (Rel. 39; Created)
DT 30-MAY-2000 (Rel. 39; Last sequence update)
DT 30-MAY-2000 (Rel. 39; Last annotation update)
DE Hypothetical protein DJ845024.5 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Graham D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RA Rhodes S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MAPE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AL022101; CAAL7880.1; -;
DR EMBL; AL049680; CAB41252.1; -;
KW Hypothetical protein.
FT NON_TER 435 435
SQ SEQUENCE 435 AA; 50539 MW; 3D19443032BBB494 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 435;
Best Local Similarity 87.5%; Pred. No. 35;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 LLPQLILL 9
I:|||||
DB 298 LVPQLILL 305

RESULT 7
YA04_HUMAN
ID YA04_HUMAN STANDARD; PRT; 500 AA.
AC O60810;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical protein DJ845024.2 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Graham D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RA Rhodes S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MAPE FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AL022101; CAAL7877.1; -;
DR EMBL; AL049681; CAB41253.1; -;
KW Hypothetical protein.
FT NON_TER 1 500
SQ SEQUENCE 500 AA; 58247 MW; 1895CD8A8F14B7C3 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 500;

. Best Local Similarity 87.5%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 2 LLPQLILL 9
I:|||||
DB 363 LVPQLILL 370

RESULT 8
AP01_MOUSE
ID AP01_MOUSE STANDARD; PRT; 653 AA.
AC Q03157;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Amyloid-like protein 1 precursor (APLP).
GN APLP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93066322; PubMed=1279693;
RA Wasco W.; Bupp K.; Magendanz M.; Gusella J.F.; Tanzi R.E.;
RT Identification of a mouse brain cDNA that encodes a protein related
RT to the Alzheimer disease-associated amyloid beta protein precursor.*;
RL Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MAY BE PROCESSED
CC IN THE GOLGI COMPLEX; HOWEVER, IT IS NOT YET CLEAR WHETHER APLP
CC IS SECRETED.

CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

DR EMBL; L04538; AAA37247.1; -;
DR PIR; A46362; A46362.
DR HSP; P05067; LMWP.

DR MGI; MGI:88046; APLP1.

DR InterPro; IPR001868; A4_APP.

DR Pfam; PF02177; A4_EXTRA; 1.

DR PRINTS; PR00203; AMYLOIDA4.

DR SMART; SM00006; A4_EXTRA; 1.

DR PROSITE; PS00319; A4_EXTRA; 1.

DR PROSITE; PS00320; A4_INTRA; 1.

KW Glycoprotein; Transmembrane; Signal.

FT SIGNAL 1 20

FT CHAIN 21 653

FT DOMAIN 21 583

FT TRANSMEM 584 606

FT DOMAIN 607 653

FT DOMAIN 643 646

FT DOMAIN 263 271

FT CARBOHYD 464 464

FT CARBOHYD 554 554

SQ SEQUENCE 653 AA; 72751 MW; 56516DC3EA40E4B0 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 653;

Best Local Similarity 77.8%; Pred. No. 52;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LLPQLILL 9
I:|||||
DB 21 LLLPLSLL 29

```

RESULT 9
CD8A_PONPY
ID CD8A_PONPY STANDARD; PRT; 198 AA.
AC P30433;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein CD8 alpha chain precursor (T-Lymphocyte
DE differentiation antigen T8/LEU-2).
GN CD8A.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ISOLATE JARI;
RX MEDLINE=92307742; PubMed=1612644;
RA Lawlor D.A., Parham P.;
RT "Structure of CD8 alpha and beta chains of the orangutan; novel
RT patterns of mRNA splicing encoding hingeless polypeptides.";
RL Immunogenetics 36:121-125(1992).
CC -!- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT
CC WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN
CC THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO
CC CLASS I MHC MOLECULES ALPHA-3 DOMAINS.
CC -!- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN
CC LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORMS HOMODIMERS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X60223; CA442784.1; *.
DR HSP; P01732; ICD8.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IGV; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell; MHC I;
FT SIGNAL.
FT CHAIN 1 21 BY SIMILARITY.
FT CHAIN 22 198 T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA
FT CHAIN 22 198 CHAIN.
FT DOMAIN 22 145 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 146 166 POTENTIAL.
FT DOMAIN 167 198 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 22 135 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 43 115 BY SIMILARITY.
FT SEQUENCE 198 AA; 22059 MW; F3EC093EADB05561 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 198;
Best Local Similarity 77.8%; Pred. No. 25;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 8 LLLPLALLL 16
|||||::|

RESULT 10
CD8A_HUMAN
ID CD8A_HUMAN STANDARD; PRT; 235 AA.
AC P01732;
DT 21-JUL-1986 (Rel. 01, Created)

```

```

DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein CD8 alpha chain precursor (T-Lymphocyte
DE differentiation antigen T8/Leu-2).
GN CD8A OR MAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85099337; PubMed=3871356;
RA Littman D.R., Thomas Y., Maddon P.J., Chess L., Axel R.;
RT "The isolation and sequence of the gene encoding T8: a molecule
RT defining functional classes of T lymphocytes.";
RL Cell 40:237-246(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86103103; PubMed=3936473;
RA Parnes J.R., Sizer K.C., Sukhatme V.P., Hunkapiller T.;
RT "Structure of Leu-2/T8 as deduced from the sequence of a cDNA clone.";
RL Behring Inst. Mitt. 77:48-55(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85124610; PubMed=3918796;
RA Sukhatme V.P., Sizer K.C., Vollmer A.C., Hunkapiller T.,
RA Parnes J.R.;
RT "The T cell differentiation antigen Leu-2/T8 is homologous to
RT immunoglobulin and T cell receptor variable regions.";
RL Cell 40:591-597(1985).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=90035142; PubMed=2509342;
RA Nakayama K.-I., Tokito S., Okumura K., Nakauchi H.;
RT "Structure and expression of the gene encoding CD8 alpha chain (Leu-
RT 2/T8).";
RL Immunogenetics 30:393-397(1989).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=89215302; PubMed=2496167;
RA Norment A.M., Lonberg N., Lacy E., Littman D.R.;
RT "Alternatively spliced mRNA encodes a secreted form of human CD8
RT alpha. Characterization of the human CD8 alpha gene.";
RL J. Immunol. 142:3312-3319(1989).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 22-135.
RX MEDLINE=92191292; PubMed=1547508;
RA Leahy D.J., Axel R., Hendrickson W.A.;
RT "Crystal structure of a soluble form of the human T cell coreceptor
RT CD8 at 2.6-A resolution.";
RL Cell 68:1145-1162(1992).
CC -!- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT
CC WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN
CC THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO
CC CLASS I MHC MOLECULES ALPHA-3 DOMAINS.
CC -!- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN
CC LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORMS HOMODIMERS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- ALTERNATIVE PRODUCTS: VARIOUS PATTERNS OF DIFFERENTIAL SPLICING
CC OF CD8 ALPHA TRANSCRIPTS INVOLVE EXCISION OF THE TRANSMEMBRANE OR
CC CYTOPLASMIC DOMAINS.
CC -!- PTM: ALL OF THE FIVE MOST CARBOXYL-TERMINAL CYSTEINES ARE USED TO
CC FORM INTER-CHAIN DISULFIDE BONDS IN DIMERS AND HIGHER MULTIMERS.
CC WHILE THE FOUR AMINO-TERMINAL CYSTEINES ARE NOT (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -!- DATABASE: NAME=PROV; NOTE=CD guide CD8a entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd8alpha.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

```
DR EMBL; M26315; AAA79217.1; -.
DR EMBL; M26313; AAA79217.1; JOINED.
DR EMBL; M26314; AAA79217.1; JOINED.
DR EMBL; M12824; AAA61133.1; -.
DR EMBL; M12828; AAB04637.1; -.
DR EMBL; M27161; AAA59674.1; -.
DR PIR; A01999; RWHU78.
DR PIR; A22824; A22824.
DR PIR; JP0105; JP0105.
DR PIR; A30604; A30604.
DR PIR; A45888; A45888.
DR PDB; 1CD8; 31-JAN-94.
DR MIM; 186910; -.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; ig; 1.
DR SMART; SM00406; IGV; 1.
DR Immunoglobulin domain; Transmembrane; Glycoprotein; Phosphorylation;
KW T-cell; MHC I; Signal; 3D-structure; Alternative splicing.
FT SIGNAL 1 21
FT CHAIN 22 235
FT T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA
FT CHAIN.
FT DOMAIN 22 182
FT TRANSMEM 183 203
FT POTENTIAL.
FT DOMAIN 204 235
FT CYTOPLASMIC (POTENTIAL).
FT DISULFID 43 115
FT STRAND 24 27
FT TURN 35 36
FT TURN 39 45
FT STRAND 54 59
FT STRAND 68 73
FT STRAND 79 80
FT TURN 82 83
FT TURN 86 88
FT STRAND 89 94
FT TURN 95 96
FT STRAND 97 102
FT HELIX 107 109
FT STRAND 111 119
FT TURN 120 121
FT STRAND 122 125
FT STRAND 129 131
SQ SEQUENCE 235 AA; 25729 MW; FCCA29BAA73726BB CRC64;

Query Match 80.0%; Score 32; DB 1; Length 235;
Best Local Similarity 77.8%; Pred. No. 30;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
    ||||| :||
Db 8 LLLPLALLL 16

RESULT 11
MABA_RAT
ID MABA_RAT STANDARD; PRT; 238 AA.
AC P19999;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 13-JUL-1999 (Rel. 38, Last annotation update)
DE Mannose-binding protein A precursor (MBP-A) (Mannan-binding protein).
GN MBP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE=Liver;
RX MEDLINE=86196130; PubMed=3009480;
RA Drickamer K., Dordal M.S., Reynolds L.;
RT "Mannose-binding proteins isolated from rat liver contain
RT carbohydrate-recognition domains linked to collagenous tails.
RT Complete primary structures and homology with pulmonary surfactant
RT apoprotein.";
RL J. Biol. Chem. 261:6878-6887(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=87137502; PubMed=3029088;
RA Drickamer K., McCreary V.;
RT "Exon structure of a mannose-binding protein gene reflects its
RT evolutionary relationship to the asialoglycoprotein receptor and
RT nonfibrillar collagens.";
RL J. Biol. Chem. 262:2582-2589(1987).
RN [3]
RP SEQUENCE OF 18-42.
RX MEDLINE=8722358; PubMed=3584121;
RA Ikeda K., Sannoh T., Kawasaki N., Kawasaki T., Yamashina I.;
RT "Serum lectin with known structure activates complement through the
RT classical pathway.";
RL J. Biol. Chem. 262:7451-7454(1987).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 124-238.
RX MEDLINE=92086855; PubMed=1721241;
RA Weis W.I., Kahn R., Fourme R., Drickamer K., Hendrickson W.A.;
RT "Structure of the calcium-dependent lectin domain from a rat mannose-
RT binding protein determined by MAD phasing.";
RL Science 254:1608-1615(1991).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.7 ANGSTROMS) OF 124-238.
RX MEDLINE=93063338; PubMed=1436090;
RA Weis W.I., Drickamer K., Hendrickson W.A.;
RT "Structure of a C-type mannose-binding protein complexed with an
RT oligosaccharide.";
RL Nature 360:127-134(1992).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 84-238.
RX MEDLINE=95219384; PubMed=7704532;
RA Weis W.I., Drickamer K.;
RT "Trimeric structure of a C-type mannose-binding protein.";
RL Structure 2:1227-1240(1994).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 90-238.
RX MEDLINE=99119227; PubMed=9922165;
RA Ng K.K.-S., Park-Snyder S., Weis W.I.;
RT "Ca2+-dependent structural changes in C-type mannose-binding
RT proteins.";
RL Biochemistry 37:17965-17976(1998).
CC -!- FUNCTION: BINDS MANNOSE AND N-ACETYLGLUCOSAMINE IN A CALCIUM-
CC DEPENDENT MANNER. IS CAPABLE OF HOST DEFENSE AGAINST PATHOGENS,
CC BY ACTIVATING THE CLASSICAL COMPLEMENT PATHWAY INDEPENDENTLY OF
CC THE ANTIBODY.
CC -!- SUBUNIT: OLIGOMERIC COMPLEX OF 6 SET OF HOMOTRIMERS.
CC -!- SUBCELLULAR LOCATION: THIS PROTEIN IS LOCATED INTRACELLULAR, MOST
CC PREDOMINANTLY IN THE ROUGH ENDOPLASMIC RETICULUM AND IN THE GOLGI
CC APPARATUS, AND LITTLE OR NO MBP IS PRESENT IN PLASMA MEMBRANES AND
CC LYOSOMES. THE MBP IS EXCLUSIVELY LOCALIZED IN THE CISTERNAL SPACE
CC OF THE ORGANELLES, PROBABLY AS A LOOSELY BOUND MEMBRANE PROTEIN.
CC -!- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M14105; AAA98781.1; -.
```

```

DR EMBL: M14104; AAA98781.1; JOINED.
DR PIR: B24791; LNRTNA.
DR PDB: 1MSB; 15-JAN-92.
DR PDB: 2MSB; 31-OCT-93.
DR PDB: 1AFA; 03-APR-96.
DR PDB: 1AFB; 03-APR-96.
DR PDB: 1AFB; 03-APR-96.
DR PDB: 1RTM; 07-FEB-95.
DR PDB: 1KMB; 12-FEB-97.
DR PDB: 1KMB; 12-FEB-97.
DR PDB: 1KMB; 12-FEB-97.
DR PDB: 1KMB; 12-FEB-97.
DR PDB: 1YTT; 10-JUN-96.
DR PDB: 1BCH; 17-JUN-98.
DR PDB: 1BCJ; 17-JUN-98.
DR PDB: 1BUU; 09-SEP-98.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE LECTIN_2; 1.
KW Lectin; Hydroxylation; Liver; Glycoprotein; Mannose-binding; Membrane;
KW Calcium; Collagen; Repeat; Signal; 3D-structure.
FT SIGNAL 1 17
FT CHAIN 18 238 MANNOSE-BINDING PROTEIN A.
FT DOMAIN 39 88 COLLAGEN-LIKE.
FT DOMAIN 143 238 C-TYPE LECTIN (SHORT FORM).
FT MOD_RES 43 43 HYDROXYLATION (POTENTIAL).
FT MOD_RES 61 61 HYDROXYLATION (POTENTIAL).
FT MOD_RES 67 67 HYDROXYLATION (POTENTIAL).
FT MOD_RES 73 73 HYDROXYLATION (POTENTIAL).
FT MOD_RES 78 78 HYDROXYLATION (POTENTIAL).
FT MOD_RES 145 234
FT DISULFID 145 234
FT DISULFID 212 226
FT CONFLICT 156 136
FT STRAND 127 136
FT HELIX 138 147
FT TURN 148 149
FT TURN 151 152
FT STRAND 158 168
FT HELIX 172 177
FT STRAND 182 183
FT TURN 186 187
FT STRAND 188 189
FT TURN 192 192
FT STRAND 198 198
FT TURN 200 201
FT TURN 207 208
FT STRAND 212 215
FT HELIX 217 219
FT STRAND 221 224
FT TURN 226 227
FT STRAND 230 237
SQ SEQUENCE 238 AA; 25308 MW; 1A927482B8A8CB3D CRC64;

Query Match 80.0%; Score 32; DB 1; Length 238;
Best Local Similarity 77.8%; Pred. No. 30;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

QY 1 LLLPLQLLL 9
DB 2 LLLPLLVLL 10
|||||:|

RESULT 12
ID CD8A_CANFA STANDARD; PRT; 239 AA.
AC P33706;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-WAR-2002 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein CD8 alpha chain precursor.

GN OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BEAGLE; TISSUE=Thymus;
RX MEDLINE=94378217; PubMed=8091416;
RA Gorman S.D., Frewin M.R., Cobbold S.P., Waldmann H.;
RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
alpha antigens.";
RL Tissue Antigens 43:184-188(1994).
CC !- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT
CC WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN
CC THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO
CC CLASS I MHC MOLECULES ALPHA-3 DOMAINS.
CC !- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN
CC LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORMS HOMODIMERS.
CC !- SUBCELLULAR LOCATION: Type I membrane protein.
CC !- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC !- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL: L14287; AAB02294.1; -.
DR HSP: P01732; LCD8.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell; MHC I;
FT SIGNAL 1 21
FT CHAIN 22 239
FT DOMAIN 22 186 T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA
FT TRANSMEM 187 210 CHAIN.
FT DOMAIN 211 239 POTENTIAL.
FT DOMAIN 25 139 CYTOPLASMIC (POTENTIAL).
FT DISULFID 46 119 IG-LIKE V-TYPE DOMAIN.
FT CARBOHYD 156 156 BY SIMILARITY.
SQ SEQUENCE 239 AA; 26036 MW; 1018579779A5CB7B CRC64;

Query Match 80.0%; Score 32; DB 1; Length 239;
Best Local Similarity 77.8%; Pred. No. 30;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
DB 8 LLLPLALLL 16
|||||:|

RESULT 13
ID MCT1_HUMAN STANDARD; PRT; 247 AA.
AC P23946; O16018;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chymase precursor (EC 3.4.21.39) (Mast cell protease I).
GN CMA1 OR CYM OR CYH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]

```

RP SEQUENCE FROM N.A.
RX MEDLINE-91302311; PubMed=2071582;
RA Caughey G.H., Zerweck E.H., Vanderslice P.;
RT "Structure, chromosomal assignment, and deduced amino acid sequence
of a human gene for mast cell chymase.";
RL J. Biol. Chem. 266:12956-12963(1991).
[2]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RX MEDLINE-91373329; PubMed=1894611;
RA Urata H., Kinoshita A., Perez D.M., Misono K.S., Bumpus F.M.,
Graham R.M., Husain A.;
RT "Cloning of the gene and cDNA for human heart chymase.";
RL J. Biol. Chem. 266:17173-17179(1991).
[3]
RP SEQUENCE OF 22-247 FROM N.A.
RX MEDLINE-93265916; PubMed=8495723;
RA Sukenaga Y., Kido H., Neki A., Enomoto M., Ishida K., Takagi K.,
Katunuma N.;
RT "Purification and molecular cloning of chymase from human tonsils.";
RL FEBS Lett. 323:119-122(1993).
[4]
RP SEQUENCE OF 26-60 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE-91264818; PubMed=2049082;
RA Jenne D.E., Tschopp J.;
RT "Angiotensin II-forming heart chymase is a mast-cell-specific
enzyme.";
RL Biochem. J. 276:567-568(1991).
[5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RX MEDLINE-98062898; PubMed=9400368;
RA McGrath M.E., Mirzadegan T., Schmidt B.F.;
RT "Crystal structure of phenylmethanesulfonyl fluoride-treated human
chymase at 1.9 A.";
RL Biochemistry 36:14318-14324(1997).
[6]
RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS).
RX MEDLINE-99134396; PubMed=9931257;
RA Pereira P.J.P., Wang Z.-M., Rubin H., Huber R., Bode W.,
Schechter N.M., Strobl S.;
RT "The 2.2-A crystal structure of human chymase in complex with
succinyl-Ala-Ala-Pro-Phe-chloromethylketone: structural explanation
for its dipeptidyl carboxypeptidase specificity.";
RL J. Mol. Biol. 286:163-173(1999).
[7]
RP ERRATUM.
RA Pereira P.J.P., Wang Z.-M., Rubin H., Huber R., Bode W.,
Schechter N.M., Strobl S.;
RL J. Mol. Biol. 286:817-817(1999).
CC -!- FUNCTION: MAJOR SECRETED PROTEASE OF MAST CELLS WITH SUSPECTED
ROLES IN VASOACTIVE PEPTIDE GENERATION, EXTRACELLULAR MATRIX
DEGRADATION, AND REGULATION OF GLAND SECRETION.
CC -!- CATALYTIC ACTIVITY: SPECIFICITY SIMILAR TO CHYMOTRYPSIN.
CC -!- SUBCELLULAR LOCATION: MAST CELL GRANULES.
CC -!- TISSUE SPECIFICITY: MAST CELLS IN LUNG, HEART, SKIN, AND PLACENTA.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPsin FAMILY. GRANZYME SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@sib-sib.ch).

CC EMBL; M691137; AAA52021.1; -
DR EMBL; M64269; AAA52020.1; -
DR EMBL; M691136; AAA52019.1; -
DR EMBL; X59072; CAA41796.1; -
DR EMBL; S61334; AAB26828.1; -
DR PIR; A40967; KYHUCM.

DR PDB; 1KLT; 13-JAN-99.
DR PDB; 1PJP; 02-MAR-99.
DR MEROPS; S01.140; -
DR MIM; 118938;
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001354; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW 3D-structure; Polymorphism.
FT SIGNAL 1 19
FT PROPEP 20 21 ACTIVATION PEPTIDE.
FT CHAIN 22 247 CHYMASE.
FT ACT_SITE 66 66
FT ACT_SITE 110 110
FT ACT_SITE 203 203
FT DISULFID 51 67
FT DISULFID 144 209
FT DISULFID 175 188
FT CARBOHYD 80 80
FT CARBOHYD 103 103
FT VARIANT 46 46
FT VARIANT 66 66
FT CONFLICT 28 28
FT SEQUENCE 247 AA; 27325 MW; DC146A049ED6B00 CRC64;
SQ
Query Match 80.0%; Score 32; DB 1; Length 247;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLLPLQLL 9
DB 2 LLLPLPLLL 10
RESULT 14
MCT1_MACFA
ID MCT1_MACFA STANDARD; PRT; 247 AA.
AC P56435;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE Chymase precursor (EC 3.4.21.39) (Mast cell protease 1).
GN CMA1.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-97400201; PubMed=9257695;
RA Takai S., Shiota N., Kobayashi S., Matsumura E., Miyazaki M.;
RT "Induction of chymase that forms angiotensin II in the monkey
atherosclerotic aorta.";
RL FEBS Lett. 412:86-90(1997).
CC -!- FUNCTION: MAJOR SECRETED PROTEASE OF MAST CELLS WITH SUSPECTED
ROLES IN VASOACTIVE PEPTIDE GENERATION, EXTRACELLULAR MATRIX
DEGRADATION, AND REGULATION OF GLAND SECRETION (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: SPECIFICITY SIMILAR TO CHYMOTRYPSIN.
CC -!- SUBCELLULAR LOCATION: MAST CELL GRANULES (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPsin FAMILY. GRANZYME SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its

CC -1- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN THE MAMMARY GLAND
 CC DURING APOPTOSIS.
 CC -1- SIMILARITY: CONTAINS 1 HEMOXIN-LIKE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
 CC METALLOPROTEASE) ALSO KNOWN AS MATRININ SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Z12604; CAA78248.1; -
 DR PIR: A44399; A44399.
 DR PDB: 1HV5; 28-MAR-01.
 DR MEROPS: M10.007; -
 DR MGD: MGI:97008; Mmp11.
 DR InterPro: IPR000585; Hemopexin.
 DR InterPro: IPR001818; Matrxin.
 DR InterPro: IPR000130; Zn_Mtpeptdse.
 DR Pfam: PF00045; hemopexin; 4.
 DR Pfam: PF00413; Peptidase_M10; 1.
 DR PRINTS: PR00138; MATRININ.
 DR SMART: SM00120; HX; 4.
 DR SMART: SM00235; ZnMC; 1.
 DR PROSITE: PS00024; HEMOXIN; 1.
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 DR PROSITE: PS00546; CYSTEINE_SWITCH; FALSE_NEG.
 KW Hydrolyase: Metalloprotease; Zinc; Zymogen; Calcium;
 KW Collagen degradation: Extracellular matrix; Signal; 3D-structure.
 FT SIGNAL 1 35 POTENTIAL.
 FT PROPEP 36 101 ACTIVATION PEPTIDE (BY SIMILARITY).
 FT CHAIN 102 492 STROMELYSIN-3.
 FT DOMAIN 295 487 HEMOXIN-LIKE.
 FT SITE 84 84 CYSTEINE SWITCH (BY SIMILARITY).
 FT METAL 219 219 ZINC (CATALYTIC).
 FT ACT_SITE 220 220 ZINC (CATALYTIC).
 FT METAL 223 223 ZINC (CATALYTIC).
 FT METAL 229 229 ZINC (CATALYTIC).
 FT DISULFID 298 484 BY SIMILARITY.
 SQ SEQUENCE 492 AA; 55441 MW; B54E260E4AB3D7C3 CRC64;
 Query Match 80.0%; Score 32; DB 1; Length 492;
 Best Local Similarity 77.8%; Pred. NO. 62;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPLQLL 9
 ||||| :||
 Db 15 LLLPLPLL 23
 RESULT 17
 AC HUMAN STANDARD; PRT; 1306 AA.
 AC P1281; 1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Angiotensin-converting enzyme, somatic isoform precursor (EC 3.4.15.1)
 DE (ACE) (Dipeptidyl carboxypeptidase I) (Kininase II) (CD143 antigen).
 GN DCPI OR DCP OR ACE.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89071703; PubMed=2849100;
 RA Soubrier F., Alhenc-Gelas F., Hubert C., Allegrini J., John M.,
 RA Tregear G., Corbol P.;
 RT "Two putative active centers in human angiotensin I-converting enzyme

RT revealed by molecular cloning.";
 RN Proc. Natl. Acad. Sci. U.S.A. 85:9386-9390(1988).
 RP SEQUENCE FROM N.A., AND VARIANTS S-261; W-561 AND S-1286.
 RX MEDLINE=99251580; PubMed=10319862;
 RA Rieder M.J., Taylor S.L., Clark A.G., Nickerson D.A.;
 RT "Sequence variation in the human angiotensin converting enzyme.";
 RL Nat. Genet. 22:59-62(1999).
 RN [3].
 RP PARTIAL SEQUENCE OF 30-46.
 RC TISSUE=Lung;
 RX MEDLINE=90110025; PubMed=2558109;
 RA Takeuchi K., Shimizu T., Ohishi N., Seyama Y., Takaku F.,
 RA Yotsumoto H.;
 RT "Purification of human lung angiotensin-converting enzyme by high-
 RT performance liquid chromatography: properties and N-terminal amino
 RL acid sequence.";
 RL J. Biochem. 106:442-445(1989).
 RN [4].
 RP ZINC-BINDING.
 RX MEDLINE=91308093; PubMed=1649623;
 RA Ehlers M.R., Riordan J.F.;
 RT "Angiotensin-converting enzyme: zinc- and inhibitor-binding
 RT stoichiometries of the somatic and testis isoforms.";
 RL Biochemistry 30:7118-7126(1991).
 CC -1- FUNCTION: CONVERTS ANGIOTENSIN I TO ANGIOTENSIN II BY RELEASE OF
 CC THE TERMINAL HIS-LEU, THIS RESULTS IN AN INCREASE OF THE
 CC VASOCONSTRICTOR ACTIVITY OF ANGIOTENSIN.
 CC -1- CATALYTIC ACTIVITY: Release of a C-terminal dipeptide,
 CC oligopeptide-[Xaa-Xbb, when Xaa is not pro, and Xbb is neither
 CC Asp nor Glu. Converts angiotensin I to angiotensin II.
 CC -1- COFACTOR: BINDS 2 ZINC IONS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: THE TESTICULAR ANGIOTENSIN-CONVERTING ENZYME
 CC IS TRANSCRIBED FROM THE SAME GENE AS THE SOMATIC ISOFORM, PROBABLY
 CC FROM AN ALTERNATIVE START SITE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M2.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD143 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd143.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04144; AAA51684.1; -
 DR EMBL: AF118569; AAD28560.1; -
 DR PIR: A31759; A31759.
 DR PIR: PQ0004; PQ0004.
 DR MIM: 106180; -
 DR InterPro: IPR001548; Peptidase_M2.
 DR InterPro: IPR000130; Zn_Mtpeptdse.
 DR Pfam: PF01401; Peptidase_M2; 2.
 DR PRINTS: PR00791; PEPTIDASEA.
 DR ProDom: PD004184; Peptidase_M2; 2.
 DR PROSITE: PS00142; ZINC_PROTEASE; 2.
 KW Hydrolyase; Metalloprotease; Carboxypeptidase; Zinc; Dipeptidase;
 KW Glycoprotein; Transmembrane; Repeat; Signal; Alternative splicing;
 KW Polymorphism. 1 29
 FT SIGNAL 30 1306 ANGIOTENSIN-CONVERTING ENZYME, SOMATIC
 FT CHAIN 30 1259 ISOFORM.
 FT DOMAIN 1260 1276 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1277 1306 POTENTIAL.
 FT DOMAIN 227 583 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 825 1181
 FT REPEAT 390 390
 FT METAL 391 391 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 394 394 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 394 394

```
FT METAL 988 988 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 989 989 ZINC 2 (CATALYTIC).
FT METAL 992 992 ZINC 2 (CATALYTIC) (BY SIMILARITY).
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 111 111 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 146 146 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 160 160 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 445 445 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 677 677 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 695 695 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 714 714 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 760 760 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 942 942 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1191 1191 N-LINKED (GLCNAC. .) (POTENTIAL).
FT VARIANT 261 261 A -> S.
FT VARIANT 561 561 R -> W.
FT VARIANT 1286 1286 /FTIG=VAR_011707.
FT VARIANT 1286 1286 /FTIG=VAR_011708.
FT CONFLICT 35 35 /FTIG=VAR_011709.
FT CONFLICT 42 42 O -> E (IN REF. 2).
FT SEQUENCE 1306 AA; 149714 MW; 1B33BCA7301A26AA CRC64;
Query Match 80.0%; Score 32; DB 1; Length 1306;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 LLLPLQILL 9
Db 12 LLLPLPLL 20
RESULT 18
ACH4_MOUSE STANDARD; PRT; 27 AA.
AC 070174;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neuronal acetylcholine receptor protein, alpha-4 chain precursor
DE (fragment).
GN CHRNA4 OR ACRA4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=98420182; PubMed=9749753;
RA Watanabe H., Zoli M., Changeux J.-P.;
RT "Promoter analysis of the neuronal nicotinic acetylcholine receptor
alpha4 gene: methylation and expression of the transgene.";
RL Eur. J. Neurosci. 10:2244-2253(1998).
CC -!- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ACHR RESPONDS BY AN
CC EXTENSIVE CHANGE IN CONFORMATION THAT AFFECTS ALL SUBUNITS AND
CC LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA
CC MEMBRANE.
CC -!- SUBUNIT: NEURONAL ACHR SEEMS TO BE COMPOSED OF TWO DIFFERENT
CC TYPE OF SUBUNITS: ALPHA AND NON-ALPHA (BETA).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
Qy 1 LLLPLQILL 9
Db 12 LLLPLPLL 20
RESULT 19
APC2_CAVPO STANDARD; PRT; 100 AA.
AC P27916;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Apolipoprotein C-II precursor (Apo-CII).
GN Apoc2.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=91154195; PubMed=1999402;
RA Andersson Y., Thelander L., Bengtsson-Olivecrona G.;
RT "Demonstration of apolipoprotein CII in guinea pigs. Functional
characteristics, cDNA sequence, and tissue expression.";
RL J. Biol. Chem. 266:4074-4080(1991).
CC -!- FUNCTION: APOC-II IS A COMPONENT OF THE VERY LOW DENSITY
CC LIPOPROTEIN (VLDL) FRACTION IN PLASMA, AND IS AN ACTIVATOR OF
CC SEVERAL TRIACYLGLYCEROL LIPASES. THE ASSOCIATION OF APOC-II WITH
CC PLASMA CHYLOMICRONS, VLDL, AND HDL IS REVERSIBLE. A FUNCTION OF
CC THE SECRETION AND CATABOLISM OF TRIGLYCERIDE-RICH LIPOPROTEINS,
CC AND CHANGES RAPIDLY.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Secreted in plasma.
CC -!- SIMILARITY: BELONGS TO THE APOC2 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M59913; AAA37031.1; -.
DR PIR; A38685; A38685.
KW Chylomicron; VLDL; Plasma; Lipid transport; Lipid degradation; Signal.
FT SIGNAL 1 25
FT CHAIN 26 100
FT SEQUENCE 100 AA; 10984 MW; BBCF8DB52FC9E9BD CRC64;
Query Match 77.5%; Score 31; DB 1; Length 100;
Best Local Similarity 77.8%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 LLLPLQILL 9
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB010002; BAA25752.1; -.
DR MGD; MGI:87888; Chrna4.
DR InterPro; IPR001175; Neur_Channel.
DR PROSITE; PS00236; NEUROTR_ION_CHANNEL; PARTIAL.
KW Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
KW Transmembrane; Multigene family.
FT SIGNAL 1 >27 POTENTIAL.
FT NON_TER 27 27
SQ SEQUENCE 27 AA; 2680 MW; FAE7C1C9664F703A CRC64;
Query Match 77.5%; Score 31; DB 1; Length 27;
Best Local Similarity 77.8%; Pred. No. 5.4;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 LLLPLQILL 9
Db 13 LLLPLLLL 21
RESULT 19
APC2_CAVPO STANDARD; PRT; 100 AA.
AC P27916;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Apolipoprotein C-II precursor (Apo-CII).
GN Apoc2.
OS Cavia porcellus (Guinea pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Hystricognathi; Caviidae; Cavia.
OX NCBI_TaxID=10141;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=91154195; PubMed=1999402;
RA Andersson Y., Thelander L., Bengtsson-Olivecrona G.;
RT "Demonstration of apolipoprotein CII in guinea pigs. Functional
characteristics, cDNA sequence, and tissue expression.";
RL J. Biol. Chem. 266:4074-4080(1991).
CC -!- FUNCTION: APOC-II IS A COMPONENT OF THE VERY LOW DENSITY
CC LIPOPROTEIN (VLDL) FRACTION IN PLASMA, AND IS AN ACTIVATOR OF
CC SEVERAL TRIACYLGLYCEROL LIPASES. THE ASSOCIATION OF APOC-II WITH
CC PLASMA CHYLOMICRONS, VLDL, AND HDL IS REVERSIBLE. A FUNCTION OF
CC THE SECRETION AND CATABOLISM OF TRIGLYCERIDE-RICH LIPOPROTEINS,
CC AND CHANGES RAPIDLY.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: Secreted in plasma.
CC -!- SIMILARITY: BELONGS TO THE APOC2 FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M59913; AAA37031.1; -.
DR PIR; A38685; A38685.
KW Chylomicron; VLDL; Plasma; Lipid transport; Lipid degradation; Signal.
FT SIGNAL 1 25
FT CHAIN 26 100
FT SEQUENCE 100 AA; 10984 MW; BBCF8DB52FC9E9BD CRC64;
Query Match 77.5%; Score 31; DB 1; Length 100;
Best Local Similarity 77.8%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 LLLPLQILL 9
```

```

Db 11 LLLPLLLL 19
||||| :||
RESULT 20
IL2_RABIT
ID IL2_RABIT STANDARD; PRT; 153 AA.
AC O77620;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
IL2.
GN Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymph node;
RA Kerr P.J., Lei S., Hardy C., Perkins H.D.;
RT "Complete cDNA sequence of rabbit interleukin-2.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M22899; AAA1427.1; -
CC PIR; A31278; A31278.
CC PIR; A45882; A45882.
CC HSSP; P01585; 3INX.
CC InterPro; IPR000779; Interleukin-2.
CC Pfam; PF00715; IL2; 1.
CC PRINTS; PR00265; INTERLEUKIN2.
CC ProDom; PD003649; Interleukin-2; 1.
CC SMART; SM00189; IL2; 1.
CC PROSITE; PS00424; INTERLEUKIN_2; 1.
CC Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
KW T-cell.
FT SIGNAL 1 20
FT CHAIN 21 155 INTERLEUKIN-2.
FT CARBOHYD 23 23 O-LINKED (GALNAC. . .) (BY SIMILARITY).
FT DISULFID 78 126 BY SIMILARITY.
FT SEQUENCE 153 AA; 17256 MW; 8173536B2DD8B86 CRC64;
KW Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
KW T-cell.
FT SIGNAL 1 20
FT CHAIN 21 155 INTERLEUKIN-2.
FT CARBOHYD 23 23 O-LINKED (GALNAC. . .) (BY SIMILARITY).
FT DISULFID 78 126 BY SIMILARITY.
FT SEQUENCE 153 AA; 17256 MW; 8173536B2DD8B86 CRC64;
Query Match 77.5%; Score 31; DB 1; Length 153;
Best Local Similarity 77.8%; Pred. No. 30;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLLPLQILL 9
||| ||:|
Db 37 LLLDLQVLL 45
RESULT 21
IL2_RAT
ID IL2_RAT STANDARD; PRT; 155 AA.
AC P17108;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
IL2.
GN Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=89339608; PubMed=2788130;
RA McKnight A.J., Mason D.W., Barclay A.N.;
RT "Sequence of rat interleukin 2 and anomalous binding of a mouse
RT interleukin 2 cDNA probe to rat MHC class II-associated invariant
RT chain mRNA.";
RL Immunogenetics 30:145-147(1989).
CC -!- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M22899; AAA1427.1; -
CC PIR; A31278; A31278.
CC PIR; A45882; A45882.
CC HSSP; P01585; 3INX.
CC InterPro; IPR000779; Interleukin-2.
CC Pfam; PF00715; IL2; 1.
CC PRINTS; PR00265; INTERLEUKIN2.
CC ProDom; PD003649; Interleukin-2; 1.
CC SMART; SM00189; IL2; 1.
CC PROSITE; PS00424; INTERLEUKIN_2; 1.
CC Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
KW T-cell.
FT SIGNAL 1 20
FT CHAIN 21 155 INTERLEUKIN-2.
FT CARBOHYD 23 23 O-LINKED (GALNAC. . .) (BY SIMILARITY).
FT DISULFID 78 126 BY SIMILARITY.
FT SEQUENCE 155 AA; 17632 MW; 67A8554A73BF30A0 CRC64;
Query Match 77.5%; Score 31; DB 1; Length 155;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLLPLQILL 9
||| ||:|
Db 37 LLLDLQVLL 45
RESULT 22
MUPM_MOUSE
ID MUPM_MOUSE STANDARD; PRT; 184 AA.
AC P04939; P97897;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Minor major urinary protein 15 precursor (Non-group 1/group 2 MUP15)
DE (MUP3).
GN MUP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/C; TISSUE=Liver;

```

```

RX MEDLINE=86135943; PubMed=3004936;
RA Clark A.J., Chave-Cox A., Ma X., Bishop J.O.;
RT "Analysis of mouse major urinary protein genes: variation between the
RT exonic sequences of group 1 genes and a comparison with an active
RT gene out with group 1 both suggest that gene conversion has occurred
RT between MUP genes.";
RN EMBO J. 4:3167-3171(1985).
RN [2]
RX SEQUENCE OF 72-184 FROM N.A.
RP MEDLINE=87257897; PubMed=3600652;
RA Shaban K., Gilmartin M., Derman E.;
RT "Nucleotide sequences of liver, lachrymal, and submaxillary gland
RT mouse major urinary protein mRNAs: mosaic structure and construction
RT of panels of gene-specific synthetic oligonucleotide probes.";
RL Mol. Cell. Biol. 7:1938-1946(1987).
CC -1- FUNCTION: BINDS PHEROMONES. THE PHEROMONES ARE RELEASED FROM
CC DRYING URINE OF MALES AND AFFECT THE SEXUAL BEHAVIOR OF FEMALES.
CC -1- TISSUE SPECIFICITY: ABUNDANT IN THE URINE OF ADULT MALE RATS
CC BUT ABSENT FROM THAT OF FEMALES.
CC -1- PTM: GLYCOSYLATED, WHILE GROUP 1 MUP PROTEINS ARE NOT.
CC -1- SIMILARITY: BELONGS TO THE LIPOCALIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X03525; CAA27228.1; -
DR EMBL; M27608; AAA39766.1; -
DR EMBL; M16359; AAA39762.1; -
DR EMBL; M16357; AAA39761.1; -
DR PIR; S10125; S10125.
DR HSSP; P02762; IMUP.
DR MGD; MGI:97235; Mup3.
DR InterPro; IPR002345; Lipocalin.
DR InterPro; IPR000566; Lipocalin_cytFABP.
DR Pfam; PF00061; Lipocalin; 1.
DR PRINTS; PR00179; LIPOCALIN.
DR PROSITE; PS00213; LIPOCALIN; 1.
KW Glycoprotein; Signal; Pheromone-binding; Transport; Lipocalin;
KW Multigene family.
FT SIGNAL 1 22
FT CHAIN 23 184 MINOR MAJOR URINARY PROTEIN 15.
FT DISULFID 86 179 BY SIMILARITY.
FT CARBOHYD 66 66 N-LINKED (GLCNAC...) (PROBABLE).
SQ SEQUENCE 184 AA; 21464 MW; EAY31827BC3972B6 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 184;
Best Local Similarity 77.8%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLILL 9
DB 3 LLLPLLLLL 11

RESULT 23
YGRS_STRCO STANDARD; PRT; 249 AA.
AC P40179;
DT 01-FEB-1995 (Rel. 31, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein SC664.38 in groES 5'/region precursor.
GN SC664.38.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae.
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]

```

```

RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Saunders D., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 121-249 FROM N.A.
RC STRAIN=A3(2) / J1501;
RX MEDLINE=94299177; PubMed=7913076;
RA Duchene A.M., Kieser H.M., Hopwood D.A., Thompson C.J., Mazodier P.;
RT "Characterization of two groEL genes in Streptomyces coelicolor
RT A3(2).";
RL Gene 144:97-101(1994).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL031317; CAA20416.1; -
DR EMBL; X75206; CAA53017.1; -
DR PIR; S37564; S37564.
DR InterPro; IPR002509; Polysac_deacet.
DR Pfam; PF01522; Polysac_deacet; 1.
KW Hypothetical protein; Signal.
FT SIGNAL 1 43 POTENTIAL.
FT CHAIN 44 249 HYPOTHETICAL PROTEIN SC664.38.
SQ SEQUENCE 249 AA; 26976 MW; 39F3188D9350A688 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 249;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLILL 9
DB 16 LLLPLLLLL 24

RESULT 24
YDJ4_SCHPO STANDARD; PRT; 338 AA.
ID YDJ4_SCHPO
AC P87052;
DT 15-JUL-1998 (Rel. 36, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Hypothetical 38.5 kDa protein C57A10.04 in chromosome I.
GN SPAC57A10.04.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Badcock K., Churcher C.M., Wood V., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 DBL-HOMOLOGY DOMAIN (DH).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z94864; CAB08167.2; -
DR InterPro; IPR000219; Rhogef.
DR SMART; SM00325; Rhogef; 1.
DR PROSITE; PS00741; DH_1; FALSE_NEG.

```

```
DR PROSITE; PSS0010; DH_2; 1.
KW Hypothetical protein.
FT DOMAIN 42 201 DH
SQ SEQUENCE 338 AA; 38463 MW; 015741B334E01106 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 338;
Best Local Similarity 77.8%; Pred. No. 66;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLILL 9
Db 150 LLLPLQLILL 158

RESULT 25
ID NCBI_HUMAN STANDARD; PRT; 461 AA.
AC Q02818; Q15838; Q9BURI;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Nucleobindin 1 precursor (CALNUC).
GN NUCB1 OR NUC.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92392352; PubMed=1520323;
RA Miura K., Titani K., Kurosawa Y., Kanai Y.;
RT "Molecular cloning of nucleobindin, a novel DNA-binding protein that
RT contains both a signal peptide and a leucine zipper structure.";
RL Biochem. Biophys. Res. Commun. 187:375-380(1992).
RN [2]
RP SEQUENCE FROM N.A., VARIANTS 13-S-DEL-14 AND ARG-399, AND REVISIONS
RP TO 299-300.
RC TISSUE=Placenta;
RX MEDLINE=96299707; PubMed=8661046;
RA Miura K., Hirai M., Kanai Y., Kurosawa Y.;
RT "Organization of the human gene for nucleobindin (NUC) and its
RT chromosomal assignment to 19q13.2-q13.4.";
RL Genomics 34:181-186(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=96026433; PubMed=7589456;
RA Wochizuki N., Hibi M., Kanai Y., Insel P.A.;
RT "Interaction of the protein nucleobindin with G alpha i2, as revealed
RT by the yeast two-hybrid system.";
RL FEBS Lett. 373:155-158(1995).
CC -1- FUNCTION: Major calcium-binding protein of the Golgi. May have a
CC role in calcium homeostasis (By similarity).
CC -1- SUBUNIT: Interacts with GNAI2.
CC -1- SUBCELLULAR LOCATION: In the cis-Golgi, tightly associated with
CC the luminal side of the membrane. A small fraction of the protein
CC may be cytoplasmic (By similarity).
CC -1- TISSUE SPECIFICITY: Expressed both in fetal and adult heart, lung,
CC liver, kidney and brain, and in adult skeletal muscle, placenta
CC and pancreas.
CC -1- MISCELLANEOUS: Discovered as DNA-binding protein in the serum of
CC lupus-prone mice.
CC -1- SIMILARITY: BELONGS TO THE NUCLEOBINDIN FAMILY.
CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96824; AAA36383.1; -
DR EMBL; U31342; AAB60431.1; -
DR EMBL; U31336; AAB60431.1; JOINED.
DR EMBL; U31337; AAB60431.1; JOINED.
DR EMBL; U31338; AAB60431.1; JOINED.
DR EMBL; U31340; AAB60431.1; JOINED.
DR EMBL; U31341; AAB60431.1; JOINED.
DR EMBL; BC002356; AAB02356.1; -
DR MIM; 601323; -
DR InterPro; IPR002048; EF-hand.
DR Pfam; PF00036; efhand; 2
DR PROSITE; PS00018; EF_HAND; 2.
KW Calcium-binding; DNA-binding; Golgi stack; Repeat; Signal;
KW Polymorphism.
FT SIGNAL 1 26 POTENTIAL.
FT CHAIN 27 461 NUCLEOBINDIN 1.
FT DNA_BIND 172 218 POTENTIAL.
FT CA_BIND 253 264 EF-HAND 1 (POTENTIAL).
FT CA_BIND 305 316 EF-HAND 2 (POTENTIAL).
FT DOMAIN 228 321 BINDS_GNAI2 (BY SIMILARITY).
FT DOMAIN 401 407 POLY-GLN.
FT VARIANT 13 14 LP -> S.
FT VARIANT 399 399 /FTID=VAR_012151.
FT VARIANT 399 399 Q -> R.
FT CONFLICT 299 300 /FTID=VAR_012152.
FT CONFLICT 385 385 HV -> QL (IN REF. 1).
FT CONFLICT 390 391 K -> Q (IN REF. 3).
FT CONFLICT 390 391 LL -> QQ (IN REF. 3).
SQ SEQUENCE 461 AA; 53821 MW; 01B7D04DF748A42A CRC64;

Query Match 77.5%; Score 31; DB 1; Length 461;
Best Local Similarity 77.8%; Pred. No. 91;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLILL 9
Db 11 LLLPLQLILL 19

RESULT 26
C72Z_ARATH
ID C72Z_ARATH STANDARD; PRT; 499 AA.
AC Q9LXM3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cytochrome P450 71B38 (EC 1.14.-.-).
GN CYP71B38 OR AT3G44250 OR T10D17_40.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=21016720; PubMed=11130713;
RA Salanoubat M., Lemcke K., Rieger M., Ansoerge W., Unseld M.,
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA De Simone V., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA Wincker P., Cholsne N., Artiguenave F., Robert C., Brottier P.,
RA Schaefer M., Cattolico L., Weissenbach J., Saurin W., Quetier F.,
RA Wurmbach E., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wiedemann R., Drzonek H., Erfle H., Jordan N., Bangert S.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordstiek G.,
RA Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
```

```

RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarsee A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Agirre A., Flores M., Liguori R., Vitale D.,
RA Manhaupt G., Haase D., Schoof H., Rudd S., Zaccaria P., Meves H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Waits A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Militscher J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Fraser C.M., Kaneko T., Nakamura Y., Sato S., Kato T., Asamizu E.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shindo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.,
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis
RT thaliana."
RL Nature 408:820-822(2000).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME P450 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AL353865; CAB88993.1; -.
DR InterPro: IPR001128; CyP450.
DR Pfam: PF00067; p450; 1.
DR PRINTS: PR00385; P450.
DR PROSITE: PS00086; CYTOCHROME_P450; 1.
KW Oxidoreductase; Monooxygenase; Transmembrane; Heme; Multigene family.
FT TRANSMEM 3 23 POTENTIAL.
FT BINDING 440 440 HEME (BY SIMILARITY).
SQ SEQUENCE 499 AA; 56779 MW; 08A2CDD3690859CA CRC64;

Query Match 77.5%; Score 31; DB 1; Length 499;
Best Local Similarity 66.7%; Pred. No. 98;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 9 LLLPLSLIL 17

RESULT 27
NDCL_HUMAN
ID NC1_HUMAN STANDARD; PRT; 592 AA.
AC Q13183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Renal sodium/dicarboxylate cotransporter (Na(+)/dicarboxylate
DE cotransporter).
DE SLC13A2 OR NADC1.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=96199379; PubMed=8967342;
RA Pajor A.M.;
RT "Molecular cloning and functional expression of a sodium-dicarboxylate
RT cotransporter from human kidney."
RL Am. J. Physiol. 270:F642-F648(1996).
CC -!- FUNCTION: COTRANSPORT OF SODIUM AND DICARBOXYLATES SUCH AS
CC SUCCINATE AND CITRATE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (potential).
CC -!- SIMILARITY: BELONGS TO THE NADC/P/PHO87 FAMILY OF TRANSPORTERS.
CC NADC SUBFAMILY.

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U26209; AAA98504.1; -.
DR MIM: 604148; -.
DR InterPro: IPR001898; Na_sulph_symp.
DR Pfam: PF00939; Na_sulph_symp; 1.
DR PROSITE: PS01271; NA_SULFATE; 1.
KW Transport; Transmembrane; Sodium transport; Symport.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 53 73 POTENTIAL.
FT TRANSMEM 86 106 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 274 294 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 450 470 POTENTIAL.
FT TRANSMEM 482 502 POTENTIAL.
FT TRANSMEM 511 531 POTENTIAL.
FT TRANSMEM 545 565 POTENTIAL.
SQ SEQUENCE 592 AA; 64410 MW; 41137D6621A0872A CRC64;

Query Match 77.5%; Score 31; DB 1; Length 592;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 23 LLLPLPLIV 31

RESULT 28
TRIL_MOUSE
ID TRIL_MOUSE STANDARD; PRT; 625 AA.
AC O35305;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Tumor necrosis factor receptor superfamily member 11A precursor
DE (Receptor activator of NF-kB) (Osteoclast differentiation factor
DE receptor) (ODFR).
GN TNFRSF11A OR RANK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal liver;
RX MEDLINE=98032977; PubMed=93671155;
RA Anderson D.M., Maraskovsky E., Billingsley W.L., Dougall W.C.,
RA Tometsko M.E., Roux E.R., Teepe M.C., Dubose R.F., Cosman D.,
RA Galibert L.;
RT "A homologue of the TNF receptor and its ligand enhance T-cell growth
RT and dendritic-cell function."
RL Nature 390:175-179(1997).
RN [2]
RP FUNCTION.
RX MEDLINE=99097247; PubMed=9878548;
RA Nakagawa N., Kinoshita M., Yamaguchi K., Shima N., Yasuda H., Yano K.,
RA Morinaga T., Higashio K.;
RT "RANK is the essential signaling receptor for osteoclast
RT differentiation factor in osteoclastogenesis."
RL Biochem. Biophys. Res. Commun. 253:395-400(1998).
CC -!- FUNCTION: RECEPTOR FOR RANK LIGAND (RANKL); ALSO KNOWN AS
CC OSTEOCLAST DIFFERENTIATION FACTOR OR ODF, ESSENTIAL FOR RANKL-

```

CC MEDIATED OSTEOCLASTOGENESIS. INVOLVED IN THE REGULATION OF
 CC INTERACTIONS BETWEEN T-CELLS AND DENDRITIC CELLS.
 CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
 CC -!- TISSUE SPECIFICITY: OBITUOUS EXPRESSION WITH HIGH LEVELS IN
 CC TRABECULAR BONE, THYMUS, SMALL INTESTINE, LUNG, BRAIN AND KIDNEY.
 CC WEAKLY EXPRESSED IN SPLEEN AND BONE MARROW.
 CC -!- SIMILARITY: CONTAINS 4 TNFR-CYS REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AF019046; AAB86810.1; -;
 CC HSSP; P25942; ICDF.
 CC MGD; MGI:1314891; Tnfrsf11a.
 CC InterPro: IPR001368; TNFR_c6.
 CC Pfam; PF00020; TNFR_c6; 3.
 CC ProDom; PD000771; TNFR_c6; 1.
 CC SMART; SM00208; TNFR; 4.
 CC PROSITE; PS00652; TNFR_NGFR_1; 1.
 CC PROSITE; PS50050; TNFR_NGFR_2; 1.
 CC Receptor; Glycoprotein; Transmembrane; Repeat; Signal.
 CC SIGNAL 1 30 POTENTIAL.
 CC CHAIN 31 625 TUMOR NECROSIS FACTOR RECEPTOR
 CC DOMAIN 31 214 EXTRACELLULAR MEMBER 11A.
 CC TRANSMEM 215 234 POTENTIAL.
 CC DOMAIN 235 625 CYTOPLASMIC (POTENTIAL).
 CC REPEAT 35 69 TNFR-CYS 1.
 CC REPEAT 72 113 TNFR-CYS 2.
 CC REPEAT 115 152 TNFR-CYS 3.
 CC REPEAT 155 195 TNFR-CYS 4.
 CC DISULFID 35 47 BY SIMILARITY.
 CC DISULFID 48 61 BY SIMILARITY.
 CC DISULFID 51 69 BY SIMILARITY.
 CC DISULFID 72 87 BY SIMILARITY.
 CC DISULFID 93 113 BY SIMILARITY.
 CC DISULFID 115 128 BY SIMILARITY.
 CC DISULFID 134 152 BY SIMILARITY.
 CC CARBOHYD 106 106 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC SEQUENCE 625 AA; 56621 MM; F8C1872E99511D8E CRC64;
 CC -----
 CC Query Match 77.5%; Score 31; DB 1; Length 625;
 CC Best Local Similarity 66.7%; Pred. No. 1.2e+02;
 CC Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 CC -----
 CC QY 1 LLLPLQLLL 9
 CC Db 21 LLVPLQVTL 29
 CC -----
 CC RESULT 29
 CC ACH4_RAT STANDARD; PRT; 630 AA.
 CC AC P09483; O35769;
 CC DT 01-MAR-1989 (Rel. 10, Created)
 CC DT 30-MAY-2000 (Rel. 39, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Neuronal acetylcholine receptor protein, alpha-4 chain precursor.
 CC CHRNA4 OR ACRA4.
 CC OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 CC NCBI_TaxID=10116;
 CC [1]
 CC RN SEQUENCE FROM N.A.
 CC RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Hypothalamus, and Hippocampus;
 CC MEDLINE=87159533; PubMed=3829125;

RA Goldman D.J., Deneris E.S., Luyten W., Kochhar A., Patrick J.,
 RA Heinemann S.F.;
 RT "Members of a nicotinic acetylcholine receptor gene family are
 RT expressed in different regions of the mammalian central nervous
 RT system.";
 RL Cell 48:965-973(1987).
 RN [2]
 RP REVISIONS.
 RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Brain;
 RA Hartley M., Goldman D.J., Heinemann S.F.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Boulter J., Deneris E.S., Evans K., Heinemann S.F.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP PARTIAL SEQUENCE OF 31-47.
 RX MEDLINE=87276531; PubMed=3609304;
 RA Whiting P., Esch F., Shimasaki S., Lindstrom J.;
 RT "Neuronal nicotinic acetylcholine receptor beta-subunit is coded for
 RT by the cDNA clone alpha 4.";
 RL FEBS Lett. 219:459-463(1987).
 CC -!- FUNCTION: AFTER BINDING ACETYLCHOLINE, THE ACHR RESPONDS BY AN
 CC EXTENSIVE CHANGE IN INFORMATION THAT AFFECTS ALL SUBUNITS AND
 CC LEADS TO OPENING OF AN ION-CONDUCTING CHANNEL ACROSS THE PLASMA
 CC MEMBRANE.
 CC -!- SUBUNIT: NEURONAL ACHR SEEMS TO BE COMPOSED OF TWO DIFFERENT TYPES
 CC OF SUBUNITS: ALPHA AND NON-ALPHA (BETA). ALPHA-4 SUBUNIT CAN BE
 CC COMBINED TO BETA-2 OR BETA-4 TO GIVE RISE TO FUNCTIONAL RECEPTORS.
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA-4-1 (SHOWN HERE) AND
 CC ALPHA-4-2: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: IN VARIOUS REGIONS OF THE CENTRAL NERVOUS
 CC SYSTEM.
 CC -!- SIMILARITY: BELONGS TO THE LIGAND-GATED IONIC CHANNELS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M15682; AAA41676.1; -;
 CC EMBL; L31620; AAC97071.1; -;
 CC EMBL; AF007212; AAB64439.1; -;
 CC PIR; A26456; A26456.
 CC InterPro: IPR000188; GABAA_receptor.
 CC InterPro: IPR001175; Neur_channel.
 CC Pfam; PF02931; Neur_chan_LBD; 1.
 CC Pfam; PF02932; Neur_chan_mem; 1.
 CC PRINTS; PR00252; NRIONCHANNEL.
 CC PROSITE; PS00236; NEUROTR_ION_CHANNEL; 1.
 CC Postsynaptic membrane; Ionic channel; Glycoprotein; Signal;
 CC Transmembrane; Multigene family; Alternative splicing.
 CC SIGNAL 1 30 POTENTIAL.
 CC CHAIN 31 630 NEURONAL ACETYLCHOLINE RECEPTOR PROTEIN,
 CC ALPHA-4 CHAIN.
 CC DOMAIN 32 249 EXTRACELLULAR (POTENTIAL).
 CC TRANSMEM 250 270 POTENTIAL.
 CC TRANSMEM 279 299 POTENTIAL.
 CC TRANSMEM 313 333 POTENTIAL.
 CC DOMAIN 334 604 CYTOPLASMIC (POTENTIAL).
 CC TRANSMEM 605 625 POTENTIAL.
 CC DISULFID 163 177 BY SIMILARITY.
 CC DISULFID 227 228 ASSOCIATED WITH RECEPTOR ACTIVATION
 CC (BY SIMILARITY).
 CC CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 109 109 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (PROBABLE).
 CC VARSPIC 629 630 AC -> GMI (IN ISOFORM ALPHA-4-2).
 CC CONFLICT 1 4 MANS -> MEIGPGGA (IN REF. 1; AAA41676 AND


```
FT CONFLICT 58 58 REF. 3).
FT CONFLICT 196 196 A -> G (IN REF. 1; AAA41676 AND REF. 3).
FT CONFLICT 301 301 M -> I (IN REF. 1; AAA41676 AND REF. 3).
SQ SEQUENCE 630 AA: 70192 MW; 1DFC9B1378E5417 CRC64;

Query Match
Best Local Similarity 77.5%; Score 31; DB 1; Length 630;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
    ||||| :||
Db 13 LLLPLLELL 21

RESULT 30
APPL_HUMAN STANDARD; PRT; 650 AA.
AC P51693; O00113;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Amyloid-like protein 1 precursor (APLP).
GN APLP1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=98088960; PubMed=9428684;
RA Paliaga K., Peraus G., Kreger S., Duwrrwang U., Hesse L., Multhaup G.,
RA Masters C.L., Beyreuther K., Weidemann A.;
RT "Human amyloid precursor-like protein 1"-cDNA cloning, ectopic
RT expression in COS-7 cells and identification of soluble forms in the
RT cerebrospinal fluid.;
RL Eur. J. Biochem. 250:354-363(1997).
RN [2]
SEQUENCE FROM N.A.
RX MEDLINE=98180887; PubMed=9521588;
RA Lenkkeri U., Kestila M., Lamerdin J., McCreedy P., Adamson A.,
RA Olsen A., Trygvason K.;
RT "Structure of the human amyloid-precursor-like protein gene APLP1 at
RT 19q13.1.";
RL Hum. Genet. 102:192-196(1998).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. C-TERMINALLY
CC PROCESSED IN THE GOLGI COMPLEX AND IS THEN SECRETED.
CC -!- TISSUE SPECIFICITY: PREDOMINANTLY EXPRESSED IN BRAIN, PARTICULARLY
CC IN THE CEREBRAL CORTEX POSTSYNAPTIC DENSITY.
CC -!- PTM: N- AND O-GLYCOSYLATED.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U48437; AAB96331.1; -
DR EMBL; AD000864; AAB50173.1; -
DR HSSP; P05067; 1MWP.
DR MIM; 104775; -
DR InterPro; IPR001868; A4_APP.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR SMART; SM00006; A4_EXTRA; 1.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
KW Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 650 AMYLOID-LIKE PROTEIN 1.
```

```
FT DOMAIN 22 580 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 581 603 POTENTIAL.
FT DOMAIN 604 650 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 640 643 CLATHRIN-BINDING (POTENTIAL).
FT DOMAIN 241 247 POLY-GLU.
FT DOMAIN 264 268 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 337 337 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 461 461 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 551 551 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 48 48 P -> A (IN REF. 2).
SQ SEQUENCE 650 AA: 72202 MW; 121A034B708C67CA CRC64;

Query Match
Best Local Similarity 77.5%; Score 31; DB 1; Length 650;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
    ||||| :||
Db 22 LLLPLLELL 30

RESULT 31
PTPE_MOUSE STANDARD; PRT; 699 AA.
AC P49446; Q62134; Q62444; Q64496;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-
DE epsilon).
GN PTPRE OR PTPE.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RP STRAIN=FVB/N;
RX MEDLINE=96064677; PubMed=7592814;
RA Elson A., Leder P.;
RT "Protein-tyrosine phosphatase epsilon. An isoform specifically
RT expressed in mouse mammary tumors initiated by v-Ha-ras or neu.";
RL J. Biol. Chem. 270:26116-26122(1995).
RN [2]
SEQUENCE FROM N.A.
RP STRAIN=DBA/2;
RA Mukoyama Y.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
SEQUENCE FROM N.A.
RP STRAIN=C57BL/6; TISSUE=Brain, and Lung;
RA Hou E.W., Li S.L.;
RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
SEQUENCE OF 224-332 FROM N.A.
RP STRAIN=BALB/C; TISSUE=Brain;
RX MEDLINE=93086603; PubMed=1454056;
RA Schepens J., Zeeuwen P., Wieringa B., Hendriks W.;
RT "Identification and typing of members of the protein-tyrosine
RT phosphatase gene family expressed in mouse brain.";
RL Mol. Biol. Rep. 16:241-248(1992).
RN [5]
SEQUENCE OF 224-332 FROM N.A.
RP STRAIN=BALB/C; TISSUE=Brain;
RX MEDLINE=95134232; PubMed=7832766;
RA Hendriks W., Schepens J., Brugman C., Zeeuwen P., Wieringa B.;
RT "A novel receptor-type protein tyrosine phosphatase with a single
RT catalytic domain is specifically expressed in mouse brain.";
RL Biochem. J. 305:499-504(1995).
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> protein
CC tyrosine + phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
```



```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U35368; AAC52281.1; -
CC EMBL: D83484; BAAL1927.1; -
CC EMBL: U62387; AAB04553.1; -
CC EMBL: Z23052; CAAR0587.1; -
CC EMBL: Z23053; CAAR0588.1; -
CC HSP: P18052; LYFO.
CC MGD: MGI:97813; Ptpre.
CC InterPro: IPR000387; Tyr_phosphatase.
CC InterPro: IPR000242; Tyr_prot_phphatase.
CC Pfam: PF00102; Y_phosphatase; 2.
CC PRINTS: PR00700; PRTYPHPHTASE.
CC SMART: SM00194; PTPC; 2.
CC PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
CC PROSITE: PS00566; TYR_PHOSPHATASE_2; 2.
CC PROSITE: PS00555; TYR_PHOSPHATASE_PTP; 2.
CC Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat;
KW SIGNAL.
FT SIGNAL 1 19
FT CHAIN 20 699
FT DOMAIN 20 45
FT TRANSMEM 46 68
FT DOMAIN 69 699
FT DOMAIN 153 392
FT DOMAIN 393 699
FT DOMAIN 334 629
FT ACT_SITE 334 334
FT ACT_SITE 629 629
FT CARBOHYD 23 23
FT CARBOHYD 31 31
FT CONFLICT 500 500
FT CONFLICT 506 506
FT CONFLICT 521 522
FT CONFLICT 606 606
FT CONFLICT 699 699
FT SEQUENCE 699 AA; 80645 MW; 4D04467438017FEB CRC64;

Query Match 77.5%; Score 31; DB 1; Length 699;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPQILL 9
   ||||| :||
Db 52 LLLPLLLLL 60

RESULT 32
PTPE_HUMAN STANDARD; PRT; 700 AA.
AC P23469;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-
DE epsilon).
GN PTPRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RA MEDLINE=91006018; PubMed=2170109;
RX Krueger N.X., Streuli M., Saito H.;
RT "Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases.";
```

```
RL EMBO J. 9:3241-3252(1990).
CC -!- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2O) -> protein
CC tyrosine + phosphate.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X54134; CAA38069.1; -
CC PIR: S12053; S12053.
CC HSP: P18052; LYFO.
CC MIM: 600926; -
CC InterPro: IPR000387; Tyr_phosphatase.
CC InterPro: IPR000242; Tyr_prot_phphatase.
CC Pfam: PF00102; Y_phosphatase; 2.
CC PRINTS: PR00700; PRTYPHPHTASE.
CC SMART: SM00194; PTPC; 2.
CC PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
CC PROSITE: PS00566; TYR_PHOSPHATASE_2; 2.
CC PROSITE: PS00555; TYR_PHOSPHATASE_PTP; 2.
CC Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat;
KW SIGNAL.
FT SIGNAL 1 19
FT CHAIN 20 700
FT DOMAIN 20 46
FT TRANSMEM 47 69
FT DOMAIN 70 700
FT DOMAIN 154 393
FT DOMAIN 394 700
FT ACT_SITE 335 335
FT ACT_SITE 630 630
FT CARBOHYD 23 23
FT CARBOHYD 30 30
FT SEQUENCE 700 AA; 80641 MW; D096BCADCEA65708 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 700;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPQILL 9
   ||||| :||
Db 51 LLLPLLLLL 59

RESULT 33
CHC2_HUMAN STANDARD; PRT; 1007 AA.
AC Q9Y5I4; Q9Y5F4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protocadherin alpha C2 precursor (PCDH-alpha-C2).
GN PCDHAC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
RC TISSUE=Brain;
RX MEDLINE=99308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
RT like cell adhesion genes.";
RL Cell 97:779-790(1999).
CC -!- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
```



```

FT CARBOHYD 746 746 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 940 940 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 442 461 MISSING (IN ISOFORM 2).
FT VARSPLIC 863 878 MISSING (IN ISOFORM 3).
FT VARSPLIC 1086 1096 MISSING (IN ISOFORM 4).
FT VARSPLIC 1279 1331 MISSING (IN ISOFORM 5).
SQ SEQUENCE 1493 AA; 163159 MW; 441DE919D5E17C0E CRC64;

Query Match 77.5%; Score 31; DB 1; Length 1493;
Best Local Similarity 77.8%; Pred. No. 2.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9
Db 27 LLLPLLLLL 35

RESULT 35
PC16_HUMAN
ID PC16_HUMAN STANDARD; PRT; 3298 AA.
AC Q96J00; O15098;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Protocadherin 16 precursor (Cadherin 19) (Cadherin fibroblast 1).
GN PCDH16 OR CDH19 OR FIB1 OR KIAA1773.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21481446; PubMed=11597768;
RA Nakajima D., Nakayama M., Kikuno R., Hirose T., Ohara O.;
RT "Identification of three novel non-classical cadherin genes through
RT comprehensive analysis of large cDNAs.";
RL Brain Res. Mol. Brain Res. 94:85-95(2001).
[2]
RN SEQUENCE OF 434-570 FROM N.A.
RP MEDLINE=97342499; PubMed=9199196;
RA Matsuyoshi N., Inamura S.;
RT Biochem. Biophys. Res. Commun. 235:355-358(1997).
CC -!- FUNCTION: Calcium-dependent cell-adhesion protein (Potential).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -!- TISSUE SPECIFICITY: Fibroblast specific.
CC -!- SIMILARITY: CONTAINS 27 CADHERIN DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL; AB053446; BAB61903.1; -
CC EMBL; AB000895; BAA21133.1; -
CC MIM; 603057; -
CC HSSP; P15116; INCU.
CC InterPro; IPR002126; Cadherin.
CC Pfam; PF00028; cadherin; 1.
CC PRINTS; PR00205; CADHERIN.
CC SMART; SM00112; CA; 1.
CC PROSITE; PS00232; CADHERIN_1; 18.
CC PROSITE; PS0268; CADHERIN_2; 27.
CC Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
CC Signal; Multigene family.
CC SIGNAL 1 42 POTENTIAL.
FT CHAIN 43 3298 PROTOCADHERIN 16.
FT DOMAIN 43 2940 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2941 2961 POTENTIAL.

```

```

2962 3298 CYTOPLASMIC (POTENTIAL).
43 143 CADHERIN 1.
144 255 CADHERIN 2.
256 382 CADHERIN 3.
367 472 CADHERIN 4.
474 578 CADHERIN 5.
579 685 CADHERIN 6.
686 790 CADHERIN 7.
791 894 CADHERIN 8.
895 1000 CADHERIN 9.
1001 1111 CADHERIN 10.
1112 1211 CADHERIN 11.
1218 1324 CADHERIN 12.
1333 1436 CADHERIN 13.
1437 1546 CADHERIN 14.
1547 1649 CADHERIN 15.
1650 1751 CADHERIN 16.
1752 1855 CADHERIN 17.
1856 1960 CADHERIN 18.
1965 2068 CADHERIN 19.
2069 2171 CADHERIN 20.
2172 2277 CADHERIN 21.
2278 2376 CADHERIN 22.
2377 2482 CADHERIN 23.
2483 2602 CADHERIN 24.
2603 2706 CADHERIN 25.
2707 2813 CADHERIN 26.
2814 2933 CADHERIN 27.
217 217 N-LINKED (GLCNAC. . .) (POTENTIAL).
256 256 N-LINKED (GLCNAC. . .) (POTENTIAL).
402 402 N-LINKED (GLCNAC. . .) (POTENTIAL).
584 584 N-LINKED (GLCNAC. . .) (POTENTIAL).
1249 1249 N-LINKED (GLCNAC. . .) (POTENTIAL).
1521 1521 N-LINKED (GLCNAC. . .) (POTENTIAL).
1718 1718 N-LINKED (GLCNAC. . .) (POTENTIAL).
1996 1996 N-LINKED (GLCNAC. . .) (POTENTIAL).
2361 2361 N-LINKED (GLCNAC. . .) (POTENTIAL).
2428 2428 N-LINKED (GLCNAC. . .) (POTENTIAL).
2569 2569 N-LINKED (GLCNAC. . .) (POTENTIAL).
2761 2761 N-LINKED (GLCNAC. . .) (POTENTIAL).
2792 2792 N-LINKED (GLCNAC. . .) (POTENTIAL).
2862 2862 N-LINKED (GLCNAC. . .) (POTENTIAL).
3298 AA; 346175 MW; 6BE8D28BEF0795DB CRC64;
SQ SEQUENCE 3298 AA; 346175 MW; 6BE8D28BEF0795DB CRC64;

Query Match 77.5%; Score 31; DB 1; Length 3298;
Best Local Similarity 77.8%; Pred. No. 6.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9
Db 21 LLLPLLLLL 29

RESULT 36
PKD1_HUMAN
ID PKD1_HUMAN STANDARD; PRT; 4303 AA.
AC P98161;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Polycystin precursor (Autosomal dominant polycystic kidney disease
DE protein 1).
DE PKD1.
GN PKD1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95254638; PubMed=7736581;
RA Gluecksmann-Kuis M.A., Taylor O., Woolf E.A., Bougueleret L.,
RA Deng N., Alperin G.D., Iris F., Hawkins F., Munro C., Lakey N.,
RA Duyk G., Schneider M.C., Geng L., Zhang F., Zhao Z., Torosian S.,

```

RA Reeders S.T., Bork P., Pohlschmidt M., Loehning C., Kraus B.,
RA Nowicka U., Leung A.I.S., Frischauf A.-M.;
RT "Polycystic kidney disease: the complete structure of the PKD1 gene
RT and its protein.";
RL Cell 81:289-298(1995).
RN [2]
RP SEQUENCE OF 2769-4303 FROM N.A.
RX MEDLINE-94273192; PubMed-8004675;
RA Ward C.J., Peral B., Hughes J., Thomas S., Gamble V.,
RA MacCarthy A.B., Sloane-Stanley J., Buckle V.J., Kearney L.,
RA Higgs D.R., Ratcliffe P.J., Harris P.C., Roelfsema J.H.,
RA Sprunt L.L., Saris J.J., Dauwerse H.G., Peters D.J.M.,
RA Breuning M.H., Nellist M., Brook-Carter P.T., Maheshwar M.M.,
RA Cordeiro I., Santos H., Cabral P., Sampson J.R., Janssen B.,
RA Hesselting-Janssen A.L.W., van den Ouweland A.M.W., Eussen B.,
RA Verhoef S., Lindhout D., Halley D.J.J.;
RT "The polycystic kidney disease 1 gene encodes a 14 kb transcript and
RT lies within a duplicated region on chromosome 16. The European
RT Polycystic Kidney Disease Consortium.";
RL Cell 77:881-894(1994).
RN [3]
RP STRUCTURE BY NMR OF 275-354
RX MEDLINE-99107746; PubMed-9869186;
RA Bycroft M., Bateman A., Clarke J., Hamill S.J., Sandford R.,
RA Thomas R.L., Chothia C.;
RT "The structure of a PKD domain from polycystin-1: implications for
RT polycystic kidney disease.";
RL EMBO J. 18:297-305(1999).
RN [4]
RP VARIANT ADPKD 3748-ARG--VAL-3752 DEL, AND VARIANT ASP-3632.
RX MEDLINE-96108969; PubMed-8554072;
RA Peral B., San Millan J.L., Ong A.C.M., Gamble V., Ward C.J.,
RA Strong C., Harris P.C.;
RT "Screening the 3' region of the polycystic kidney disease 1 (PKD1)
RT gene reveals six novel mutations.";
RL Am. J. Hum. Genet. 58:86-96(1996).
RN [5]
RP VARIANTS ADPKD P-2993; R-3016 AND V-3511, AND VARIANTS M-3510 AND
RP F-4190.
RX MEDLINE-97342914; PubMed-9199561;
RA Peral B., Gamble V., Strong C., Ong A.C.M., Sloane-Stanley J.,
RA Zerres K., Winears C.G., Harris P.C.;
RT "Identification of mutations in the duplicated region of the
RT polycystic kidney disease 1 gene (PKD1) by a novel approach.";
RL Am. J. Hum. Genet. 60:1399-1410(1997).
RN [6]
RP VARIANT ALA-4058.
RX MEDLINE-97295081; PubMed-9150733;
RA Constantinides R., Xenophontos S.L., Neophytou P., Nomura S.,
RA Pierides A., Constantinou-Deltas C.D.;
RT "New amino acid polymorphism, Ala/Val4058, in exon 45 of the
RT polycystic kidney disease 1 gene: evolution of alleles.";
RL Hum. Genet. 99:644-647(1997).
RN [7]
RP VARIANTS T-2760; P-2761; V-2763; T-2764; Q-2791; T-2826; L-3008 AND
RP L-3064.
RX MEDLINE-97449169; PubMed-9285784;
RA Watnick T.J., Piontek K.B., Cordal T.M., Weber H., Gandolph M.A.,
RA Qian F., Lens X.M., Neumann H.P.H., Germino G.G.;
RT "An unusual pattern of mutation in the duplicated portion of PKD1 is
RT revealed by use of a novel strategy for mutation detection.";
RL Hum. Mol. Genet. 6:1473-1481(1997).
RN [8]
RP VARIANT ADPKD THR-3678.
RX MEDLINE-97403939; PubMed-9259200;
RA Turco A.E., Rossetti S., Bresin E., Englisch S., Corra S.,
RA Pignatti P.F.;
RT "Three novel mutations of the PKD1 gene in Italian families with
RT autosomal dominant polycystic kidney disease.";
RL Hum. Mutat. 10:164-167(1997).
RN [9]
RP VARIANT ADPKD ASP-4032, AND VARIANT VAL-4045.
RX MEDLINE-98180892; PubMed-9521593;

RA Danielis C., Maheshwar M.M., Lazarou L., Davies F., Coles G.,
RA Ravine D.;
RT "Novel and recurrent mutations in the PKD1 (polycystic kidney
RT disease) gene.";
RL Hum. Genet. 102:216-220(1998).
RN [10]
RP VARIANT ADPKD MET-3375.
RX MEDLINE-99118881; PubMed-9921908;
RA Koptides M., Constantinides R., Kyriakides G., Hadjigavriel M.,
RA Patsalis P.C., Pierides A., Deltas C.C.;
RT "Loss of heterozygosity in polycystic kidney disease with a missense
RT mutation in the repeated region of PKD1.";
RL Hum. Genet. 103:709-717(1998).
RN [11]
RP VARIANTS ADPKD L-324 AND S-845, AND VARIANTS R-1399 AND L-1786.
RX MEDLINE-99294380; PubMed-10364515;
RA Thomas R.L., McConnell R., Whittacker J., Kirkpatrick P., Bradley J.,
RA Sandford R.;
RT "Identification of mutations in the repeated part of the autosomal
RT dominant polycystic kidney disease type 1 gene, PKD1, by long-range
RT PCR.";
RL Am. J. Hum. Genet. 65:39-49(1999).
RN [12]
RP VARIANTS ADPKD P-2392 AND F-2423, AND VARIANTS R-1399; Q-2548 AND
RP R-2638.
RX PubMed-10577909;
RA Watnick T., Phakdeekitcharoen B., Johnson A., Gandolph M., Wang M.,
RA Briefel G., Klinger K.W., Kimberling W., Gabow P., Germino G.G.;
RT "Mutation detection of PKD1 identifies a novel mutation common to
RT three families with aneurysms and/or very-early-onset disease.";
RL Am. J. Hum. Genet. 65:1561-1571(1999).
RN [13]
RP VARIANTS ADPKD L-3994--F-3996 DUPL; G-4136 AND C-4154, AND VARIANTS.
RX MEDLINE-20441957; PubMed-10987650;
RA Perrichot R.A., Mercier B., Simon P.M., Whebe B., Cledes J., Ferec C.;
RT "DGGE screening of PKD1 gene reveals novel mutations in a large cohort
RT of 146 unrelated patients.";
RL Hum. Genet. 105:231-239(1999).
RN [14]
RP VARIANTS ADPKD 3748-R-V-3752 DEL AND L-4132 DEL, AND VARIANT V-4045.
RX MEDLINE-20112427; PubMed-10647901;
RA Afzal A.R., Hand M., Ternes-Pereira E., Saggar-Malik A., Taylor R.,
RA Jeffery S.;
RT "Novel mutations in the 3 region of the polycystic kidney disease 1
RT (PKD1) gene.";
RL Hum. Genet. 105:648-653(1999).
RN [15]
RP VARIANTS ADPKD PRO-4225 AND TRP-4276.
RX MEDLINE-99217041; PubMed-10200984;
RA Badenas C., Torra R., San Millan J.L., Lucero L., Mila M.,
RA Estivill X., Darnell A.;
RT "Mutational analysis within the 3' region of the PKD1 gene.";
RL Kidney Int. 55:1225-1233(1999).
RN [16]
RP VARIANTS ADPKD MET-2250 AND TRP-2329, AND VARIANTS CYS-2379; LEU-3066;
RP VAL-3139 AND LEU-3193.
RX MEDLINE-20311156; PubMed-10854095;
RA Perrichot R., Mercier B., Quere I., Carre A., Simon P., Whebe B.,
RA Cledes J., Ferec C.;
RT "Novel mutations in the duplicated region of PKD1 gene.";
RL Eur. J. Hum. Genet. 8:353-359(2000).
RN [17]
RP VARIANTS ADPKD PRO-2921 AND MET-3375, AND VARIANT LEU-3066.
RX MEDLINE-20382887; PubMed-10923040;
RA Koptides M., Mean R., Demetriou K., Constantinides R., Pierides A.,
RA Harris P.C., Deltas C.C.;
RT "Screening of the PKD1 duplicated region reveals multiple single
RT nucleotide polymorphisms and a de novo mutation in Hellenic
RT polycystic kidney disease families.";
RL Hum. Mutat. 16:176-176(2000).
RN [18]
RP VARIANTS ADPKD GLN-3719 AND PRO-3852, AND VARIANT VAL-4045.
RX MEDLINE-20514565; PubMed-11058904;

RA Aguiari G., Savelli S., Garbo M., Bozza A., Augello G., Penolazzi L.,
RA De Paoli Vitali E., La Torre C., Cappelli G., Piva R., del Senno L.;
RT "Novel splicing and missense mutations in autosomal dominant
RT polycystic kidney disease 1 (PKD1) gene: expression of mutated
RT genes."; Hum. Mutat. 16:444-445(2000).
RN [19]
RN VARIANTS ADPKD SER-1166; GLU-1956; CYS-2408 AND GLY-2442--2443 INS,
RP AND VARIANTS HTS-1995 AND ASN-2604.
RX MEDLINE-20467506; PubMed-11012875;
RA Phakdeekitcharoen B., Watnick T.J., Ahn C., Whang D.-Y., Burkhardt B.,
RA Germino G.G.; "Thirteen novel mutations of the replicated region of PKD1 in an Asian
RT population."; Kidney Int. 58:1400-1412(2000).
RL [20]
RN VARIANTS ADPKD TRP-3753 AND ASN-3815.
RX MEDLINE-20275386; PubMed-10729710;
RA Kim U.K., Jin D.K., Ahn C., Shin J.H., Lee K.B., Kim S.H., Chae J.J.,
RA Hwang D.Y., Lee J.G., Namkoong Y., Lee C.C.;
RT "Novel mutations of the PKD1 gene in Korean patients with autosomal
RT dominant polycystic kidney disease."; Mutat. Res. 432:39-45(2000).
RL [21]
RN VARIANTS ADPKD Q-13; F-75; C-139; 1992-F-T-1993 DELINS L; 2220-R--P-
RP 2224 DEL; D-2336; D-2752; 2762-ILMR-2765 DUPL; M-2768; K-2771; P-2816;
RP S-2858; 3012-T--Y-3017 DEL AND 3748-L-R-3752 DEL, AND VARIANTS S-
RP 2674; M-2708; T-2734; L-2735; C-2765; M-2782; R-2814; G-2888; I-2905;
RP D-2966 AND L-3066.
RX MEDLINE-21063179; PubMed-11115377;
RA Rossetti S., Strmecki L., Gamble V., Burton S., Sneddon V., Peral B.,
RA Roy S., Bakaloglu A., Komel R., Winearls C.G., Harris P.C.;
RT "Mutation analysis of the entire PKD1 gene: genetic and diagnostic
RT implications."; Am. J. Hum. Genet. 68:46-63(2001).
RL [22]
RN Query Match 77.5%; Score 31; DB 1; Length 4303;
Best Local Similarity 55.6%; Pred. No. 8.4e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
:||||:
Db 2153 VLPPLQVLM 2161

RESULT 37
PF4L_PIG STANDARD; PRT; 119 AA.
AC P43030;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Platelet basic protein precursor (PBP).
GN SCYB7 OR PPBP.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RN SEQUENCE FROM N.A., AND SEQUENCE OF 40-119.
RC Tissue-platelet;
RX MEDLINE-94229068; PubMed-7513641;
RA Power C.A., Proudfoot A.E.I., Magnat E., Bacon K.B., Wells T.N.C.;
RT "Molecular cloning and characterisation of a neutrophil chemotactic
RT protein from porcine platelets."; Eur. J. Biochem. 221:713-719(1994).
RL C-X-C (CHEMOKINE CXCL).
CC -1- FUNCTION: CHEMOATTRACTANT FACTOR FOR NEUTROPHILS.
CC -1- MASS SPECTROMETRY: MW=8997.5; METHOD=Electrospray; RANGE=40-119.
CC -1- SIMILARITY: BELONGS TO THE INTERCRINE ALPHA FAMILY (SMALL CYTOKINE
CC C-X-C) (CHEMOKINE CXCL).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X77935; CAA54907.1; -.
DR HSP: P02775; INAP.
DR InterPro: IPR001811; Chemokine_IL8.
DR InterPro: IPR001089; Small_cytokine_CXC.
DR Pfam: PF00048; IL8; 1.
DR PRINTS: PR00437; SMALLCYTKXC.
DR SMART: SM00199; SCY; 1.
DR PROSITE: PS00471; SMALL_CYTOKINES_CXC; 1.
KW Cytokine; Growth factor; Chemotaxis; Mitogen; Platelet; Signal.
FT SIGNAL 1 33 PROBABLE.
FT PROPEP 34 39
FT CHAIN 40 119 PLATELET BASIC PROTEIN.
FT DISULFID 54 80 BY SIMILARITY.
FT DISULFID 56 96 BY SIMILARITY.
SQ SEQUENCE 119 AA; 12615 MW; 79E1D409CDD06B32 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 119;
Best Local Similarity 66.7%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
:||||:
Db 22 VLLPLSLL 30

RESULT 38
EXSB_PSEAE STANDARD; PRT; 137 AA.
AC P26994;
DT 01-AUG-1992 (Rel. 23, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Exoenzyme S synthesis protein B.
GN EXSB OR PA1712.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE-92011420; PubMed-1655713;
RA Frank D.W., Iglewski B.H.;
RT "Cloning and sequence analysis of a trans-regulatory locus required
RT for exoenzyme S synthesis in Pseudomonas aeruginosa."; J.
RL J. Bacteriol. 173:6460-6468(1991).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PAO1;
RX MEDLINE-20437337; PubMed-10984043;
RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PAO1, an
RT opportunistic pathogen."; Nature 406:959-964(2000).
RL Nature 406:959-964(2000).
CC -1- SIMILARITY: TO Y ENTEROCOLITICA VTRB.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: M64975; AAA25815.1; -
DR EMBL: AB004597; AAG05101.1; -
DR PIR: B41047; B41047.
KW Complete proteome.
FT CONFLICT 62 62 R -> W (IN REF. 1).
SQ SEQUENCE 137 AA; 14996 MW; F7367E7AB301C4EA CRC64;

Query Match 75.0%; Score 30; DB 1; Length 137;
Best Local Similarity 66.7%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 LLLPQLILL 9
DB 1 MLLPLALLL 9

RESULT 39

IL2_PIG STANDARD; PRT; 154 AA.
AC P26891;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
GN IL2.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
RX MEDLINE=91274360; PubMed=2054386;
RA Goodall J.C., Emery D.C., Bailey M., English L.S., Hall L.;
RT "cDNA cloning of porcine interleukin 2 by polymerase chain reaction.";
RL Biochim. Biophys. Acta 1089:257-258(1991).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=T-cell;
RA Lefevre F.;
RL Submitted (MAY-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Iwata H., Hasegawa A., Yamamoto M., Oida T., Endo Y., Inoue T.;
RT "Structure of the porcine chromosomal interleukin-2 gene.";
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: X56750; CAA40071.1; -
DR EMBL: X58428; CAA41330.1; -
DR EMBL: AB041935; BAB16110.1; -
DR PIR: S15473; S15473.
DR PIR: S16241; S16241.
DR HSP: P01585; 3INK.
DR InterPro: IPR000779; Interleukin-2.
DR Pfam: PF00715; IL2; 1.

DR PRINTS; PR00265; INTERLEUKIN2.
DR ProDom; PD003649; Interleukin-2; 1.
DR SMART; SM00189; IL2; 1.
DR PROSITE; PS00424; INTERLEUKIN_2; 1.
KW Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
KW T-cell.
FT SIGNAL 1 20 BY SIMILARITY.
FT CHAIN 21 154 INTERLEUKIN-2.
FT CARBOHYD 23 23 O-LINKED (GALNAC. . .) (BY SIMILARITY).
FT DISULFID 78 126 BY SIMILARITY.
SQ SEQUENCE 154 AA; 17401 MW; F3B95E43D4A3D3E1 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 154;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LLLPQLILL 9
DB 37 LLLDLQLLL 45

RESULT 40

IL2_BOVIN STANDARD; PRT; 155 AA.
AC P05016;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
GN IL2 OR IL-2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86205869; PubMed=3517854;
RA Cerretti D.P., McKerghan K., Larsen A., Cantrell M.A., Anderson D.,
RA Gillis S., Cosman D., Baker P.E.;
RT "Cloning, sequence, and expression of bovine interleukin 2.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3223-3227(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86205870; PubMed=3486415;
RA Reeves R., Spies A.G., Nissen M.S., Buck C.D., Weinberg A.D.,
RA Barr P.J., Magnuson N.S., Magnuson J.A.;
RT "Molecular cloning of a functional bovine interleukin 2 cDNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:3228-3232(1986).
RN [3]
RP SEQUENCE OF 1-22 FROM N.A.
RC TISSUE=Thymus;
RA Anikveva N.N., Vinogradova T.V., Votoshin O.N.;
RL Submitted (DEC-1989) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

DR EMBL: M12791; AAA30586.1; -
DR EMBL: M13204; AAA21143.1; ALT_INIT.
DR EMBL: X17201; CAA35062.1; -

DR EMBL; X52687; CAA36912.1; -
 DR HSP; P01585; 3INK.
 DR InterPro: IPR000779; Interleukin-2.
 DR Pfam: PF00715; IL2; 1.
 DR PRINTS; PR00265; INTERLEUKIN2.
 DR ProDom; PD003649; Interleukin-2; 1.
 DR SMART; SM00189; IL2; 1.
 DR PROSITE; PS00424; INTERLEUKIN_2; 1.
 KW Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
 KW T-cell.
 FT SIGNAL 1 20
 FT CHAIN 21 155 INTERLEUKIN-2.
 FT DISULFID 79 127 BY SIMILARITY.
 FT CARBOHYD 23 23 O-LINKED (GALNAc...) (BY SIMILARITY).
 FT CONFLICT 66 66 V -> A (IN REF. 2).
 SQ SEQUENCE 155 AA; 17627 MW; 816667DFA052EDF CRC64;
 Query Match 75.0%; Score 30; DB 1; Length 155;
 Best Local Similarity 77.8%; Pred. No. 48;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPLQLILL 9
 Db 37 LLLDLQLLLL 45
 RESULT 41
 IL2_CANFA STANDARD; PRT; 155 AA.
 AC Q29416; Q28249;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
 IL2.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP STRAIN=XBRED21/12/93; TISSUE=Lymph node;
 RX MEDLINE=95337423; PubMed=7612930;
 RA Dunham S.P., Argyle D.J., Onions D.E.;
 RT "The isolation and sequence of canine interleukin-2.";
 RL DNA Seq. 5:177-180(1995).
 RN [2]
 RC SEQUENCE FROM N.A.
 RP MEDLINE=96016696; PubMed=8571541;
 RX Somborg R.L., Pullen R.P., Casal M.L., Patterson D.F., Felsburg P.J.,
 RA Henthorn P.S.;
 RT "A single nucleotide insertion in the canine interleukin-2 receptor
 gamma chain results in X-linked severe combined immunodeficiency
 disease.";
 RL Vet. Immunol. Immunopathol. 47:203-213(1995).
 RN [3]
 RC SEQUENCE FROM N.A.
 RP STRAIN=BEAGLE; TISSUE=Spleen;
 RX MEDLINE=95347614; PubMed=7622066;
 RA Knapp D.W., Williams J.S., Andrisani O.M.;
 RT "Cloning of the canine interleukin-2-encoding cDNA.";
 RL Gene 159:281-282(1995).
 CC -1- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
 CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL
 CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
 CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
 CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; D30710; BAA06378.1; -
 DR EMBL; U28141; AAA68969.1; -
 DR EMBL; U11689; AAA75360.1; -
 DR HSP; P01585; 3INK.
 DR InterPro: IPR000779; Interleukin-2.
 DR Pfam: PF00715; IL2; 1.
 DR PRINTS; PR00265; INTERLEUKIN2.
 DR ProDom; PD003649; Interleukin-2; 1.
 DR SMART; SM00189; IL2; 1.
 DR PROSITE; PS00424; INTERLEUKIN_2; 1.
 KW Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
 KW T-cell.
 FT SIGNAL 1 20 BY SIMILARITY.
 FT CHAIN 21 155 INTERLEUKIN-2.
 FT CARBOHYD 24 24 O-LINKED (GALNAc...) (BY SIMILARITY).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC...) (POTENTIAL).
 FT DISULFID 79 127 BY SIMILARITY.
 FT CONFLICT 4 4 M -> I (IN REF. 3).
 FT CONFLICT 37 37 Q -> R (IN REF. 3).
 FT CONFLICT 151 151 F -> Y (IN REF. 3).
 FT CONFLICT 154 154 L -> M (IN REF. 3).
 SQ SEQUENCE 155 AA; 17668 MW; DL23E486B7F4AC1D CRC64;
 Query Match 75.0%; Score 30; DB 1; Length 155;
 Best Local Similarity 77.8%; Pred. No. 48;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPLQLILL 9
 Db 38 LLLDLQLLLL 46
 RESULT 42
 IL2_CAPHI STANDARD; PRT; 155 AA.
 ID IL2_CAPHI
 AC P36835; P79156;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
 IL2.
 OS Capra hircus (Goat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Caprinae; Capra.
 OX NCBI_TaxID=9925;
 RN [1]
 RC SEQUENCE FROM N.A.
 RP TISSUE=Blood;
 RC Rimstad E.;
 RA Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RC SEQUENCE FROM N.A.
 RP Beyer J.C., Cheevers W.P.;
 RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
 CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL
 CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
 CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
 CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
-----
CC EMBL; X76063; CAA53664.1; -.
CC EMBL; U34274; AAB38527.1; -.
CC DR PIR; S38662; S38662.
CC DR HSSP; P01585; 3INK.
CC DR InterPro; IPR000779; Interleukin-2.
CC DR PRINTS; PR00265; INTERLEUKIN2.
CC DR ProDom; PD003649; Interleukin-2; 1.
CC DR SMART; SM00189; IL2; 1.
CC DR PROSITE; PS00424; INTERLEUKIN_2; 1.
CC DR Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
KW T-cell.
FT CHAIN 1 20 BY SIMILARITY.
FT CHAIN 21 155 INTERLEUKIN-2.
FT CARBOHYD 23 23 O-LINKED (GALNAC. . .) (BY SIMILARITY).
FT DISULFID 79 127 BY SIMILARITY.
FT CONFLICT 3 5 RMO -> QIP (IN REF. 2).
FT CONFLICT 22 22 P -> T (IN REF. 2).
FT CONFLICT 30 30 T -> P (IN REF. 2).
FT CONFLICT 51 51 L -> P (IN REF. 2).
FT CONFLICT 71 71 D -> A (IN REF. 2).
FT CONFLICT 89 89 D -> E (IN REF. 2).
FT CONFLICT 99 99 R -> L (IN REF. 2).
FT CONFLICT 107 113 YMASLKG -> SMDNIKR (IN REF. 2).
FT CONFLICT 140 140 Q -> L (IN REF. 2).
FT CONFLICT 144 144 T -> I (IN REF. 2).
FT CONFLICT 154 154 L -> M (IN REF. 2).
FT CONFLICT 155 155 AA; 17703 MW; 90022DFB6AF78DE CRC64;
SQ SEQUENCE 155 AA; 17703 MW; 90022DFB6AF78DE CRC64;

Query Match 75.0%; Score 30; DB 1; Length 155;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 37 LLLDLQLLL 45

RESULT 43
IL2_SHEEP
ID IL2_SHEEP STANDARD; PRT; 155 AA.
AC P19114;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
GN Il2.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91016933; PubMed=2216781;
RA Goodall J.C., Emery D.C., Perry A.C.F., English L.S., Hall L.;
RT "cDNA cloning of ovine interleukin 2 by PCR.";
RL Nucleic Acids Res. 18:5883-5883(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91088336; PubMed=2263496;
RA Seow H.F., Rothel J.S., Radford A.J., Wood P.R.;
RT "The molecular cloning of ovine interleukin 2 gene by the polymerase
chain reaction.";
RL Nucleic Acids Res. 18:7175-7175(1990).
RN [3]
RP SEQUENCE OF 21-153 FROM N.A.
RX MEDLINE=95367627; PubMed=7543777;
RA Bujdoso R., Williamson M.L., Roy D., Hunt P., Blacklaws B., Sargan D.,
RA McConnell I.;
```

```
RT RT "Molecular cloning and expression of DNA encoding ovine Interleukin
2.";
RL Cytokine 7:223-231(1995).
CC -!- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

```
-----
CC EMBL; X53934; CAA37881.1; -.
CC EMBL; X55641; CAA39165.1; -.
CC EMBL; X60148; CAA42722.1; -.
CC EMBL; A19169; CAA01448.1; -.
CC DR PIR; S11488; S11488.
CC DR PIR; S13102; S13102.
CC DR PIR; S15517; S15517.
CC DR HSSP; P01585; 3INK.
CC DR InterPro; IPR000779; Interleukin-2.
CC DR Pfam; PF00715; IL2; 1.
CC DR PRINTS; PR00265; INTERLEUKIN2.
CC DR ProDom; PD003649; Interleukin-2; 1.
CC DR SMART; SM00189; IL2; 1.
CC DR PROSITE; PS00424; INTERLEUKIN_2; 1.
CC DR Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
KW T-cell.
FT SIGNAL 1 20
FT CHAIN 21 155 INTERLEUKIN-2.
FT CARBOHYD 23 23 O-LINKED (GALNAC. . .) (BY SIMILARITY).
FT DISULFID 79 127 BY SIMILARITY.
FT CONFLICT 6 6 P -> L (IN REF. 2).
SQ SEQUENCE 155 AA; 17661 MW; 3EAC6D67667494D CRC64;

Query Match 75.0%; Score 30; DB 1; Length 155;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 37 LLLDLQLLL 45

RESULT 44
IL2_CEREL
ID IL2_CEREL STANDARD; PRT; 162 AA.
AC P51747;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
GN IL2.
OS Cervus elaphus (Red deer).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Cervidae;
OC Cervidae; Cervinae; Cervus.
OX NCBI_TaxID=9860;
RN [1]
RP SEQUENCE FROM N.A.
RA Lockhart E.A.;
RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
CC MITOGENIC STIMULATION, THIS PROTEIN IS REQUIRED FOR T-CELL
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
```



```

CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS (BY
CC SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14682; AAA61733.1; -.
CC HSP: P01585; 3INK.
CC InterPro: IPR000779; Interleukin-2.
CC Pfam: PF00715; IL2; 1.
CC PRINTS: PR00285; INTERLEUKIN2.
CC ProDom: PD003649; Interleukin-2; 1.
CC SMART: SM00189; IL2; 1.
CC PROSITE: PS00424; INTERLEUKIN_2; 1.
CC Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
CC T-cell.
CC SIGNAL 1 20 BY SIMILARITY.
CC CHAIN 21 162 INTERLEUKIN-2.
CC DISULFID 79 134 BY SIMILARITY.
CC CARBOHYD 23 23 O-LINKED (GLNAC. . .) (BY SIMILARITY).
CC CARBOHYD 70 70 N-LINKED (GLNAC. . .) (POTENTIAL).
CC SEQUENCE 162 AA; 18389 MW; 6E57299B1B6E799 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 162;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 37 LLLDLQILL 45

RESULT 45
CUTI_BOTCI STANDARD; PRT; 202 AA.
AC Q00298;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cutinase precursor (EC 3.1.1.74).
GN CUTA.
OS Botrytis cinerea (Botryotinia fuckeliana).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
OC Helotiales; Sclerotiniaceae; Botryotinia.
OX NCBI_TaxID=40559;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SAS56;
RX MEDLINE=97155570; PubMed=9002269;
RA van der Vlugt-Bergmans C.J.B., Wagemakers L.C.A.M., van Kan J.A.L.;
RT "Cloning and expression of the cutinase A gene of Botrytis cinerea.";
RL Mol. Plant Microbe Interact. 10:21-29(1997).
CC -|- FUNCTION: CATALYZE THE HYDROLYSIS OF CUTIN, A POLYESTER THAT FORMS
CC THE STRUCTURE OF PLANT CUTICLE. ALLOWS PATHOGENIC FUNGI TO
CC PENETRATE THROUGH THE CUTICULAR BARRIER INTO THE HOST PLANT DURING
CC THE INITIAL STAGE OF THE FUNGAL INFECTION (By similarity).
CC -|- CATALYTIC ACTIVITY: Cutin + H(2)O -> cutin monomers.
CC -|- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -|- PTM: THE 2 DISULFIDE BONDS PLAY A CRITICAL ROLE IN HOLDING THE
CC CATALYTIC RESIDUES IN JUXTA-POSITION; REDUCTION OF THE DISULFIDE
CC BRIDGES RESULTS IN THE COMPLETE INACTIVATION OF THE ENZYME (BY
CC SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE CUTINASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z69264; CAA93255.1; -.
CC HSP: P00590; 1X2J.
CC InterPro: IPR000675; Cutinase.
CC Pfam: PF01083; Cutinase; 1.
CC PROSITE: PS00155; CUTINASE_1; 1.
CC PROSITE: PS00931; CUTINASE_2; 1.
CC Hydrolase; Serine esterase; Signal.
CC SIGNAL 1 20 POTENTIAL.
CC CHAIN 21 202 CUTINASE.
CC DISULFID 31 172 BY SIMILARITY.
CC DISULFID 106 165 BY SIMILARITY.
CC ACT_SITE 117 117 BY SIMILARITY.
CC ACT_SITE 169 169 BY SIMILARITY.
CC ACT_SITE 182 182 BY SIMILARITY.
CC SEQUENCE 202 AA; 20253 MW; 031D64725A23DBE7 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 202;
Best Local Similarity 75.0%; Pred. No. 62;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 8
DB 12 LLLPLSVL 19

Search completed: November 6, 2002, 12:09:07
Job time : 8.33333 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 11 Seconds
(without alignments)
78.619 Million cell updates/sec

Title: US-09-905-083-33
Perfect score: 40
Sequence: 1 LLLPLQILL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	253	2 A53968	serine proteinase
2	36	90.0	146	2 H75201	hypothetical prote
3	34	85.0	277	2 H84314	cytochrome aa3 con
4	34	85.0	369	2 S33603	surfactant protein
5	34	85.0	371	1 JN0450	conglutinin precu
6	34	85.0	371	2 I45878	conglutinin - bovi
7	34	85.0	754	2 AE0614	probable competenc
8	34	85.0	845	2 D97163	cation transport p
9	33	82.5	196	2 G65039	hypothetical prote
10	33	82.5	282	2 E70890	hypothetical prote
11	33	82.5	370	2 AB3334	daunorubicin resis
12	33	82.5	398	2 C91063	hypothetical prote
13	33	82.5	413	2 AC0834	probable membrane
14	33	82.5	426	2 C83103	hypothetical prote
15	33	82.5	470	2 A90083	hypothetical prote
16	33	82.5	475	1 A69149	O-antigen transpor
17	33	82.5	653	2 A46362	amyloid precursor-
18	32	80.0	198	2 S25656	T-cell surface gly
19	32	80.0	235	1 RWHTU8	mannose-binding le
20	32	80.0	238	1 LNRTMA	chymase (EC 3.4.21
21	32	80.0	247	1 KYHUCM	type III secretion
22	32	80.0	249	2 B85941	hypothetical prote
23	32	80.0	249	2 C97402	probable acyltrans
24	32	80.0	264	2 AC2620	1-acyl-sn-glycerol
25	32	80.0	264	2 AC2620	tax-responsive ele
26	32	80.0	266	2 JC7300	hepatocarcinogenes
27	32	80.0	267	2 JC4857	stromelysin 3 (EC
28	32	80.0	491	2 JC6197	stromelysin 3 (EC
29	32	80.0	492	2 A44399	

30	32	80.0	543	2 AI2088	Na+/H+-exchanging
31	32	80.0	1306	1 A31759	peptidyl-dipeptida
32	31	77.5	100	2 A38685	apolipoprotein C-I
33	31	77.5	155	2 A31278	interleukin-2 prec
34	31	77.5	159	2 G75555	conserved hypothet
35	31	77.5	184	2 S10125	alpha-2u-globulin
36	31	77.5	189	2 T43766	hypothetical prote
37	31	77.5	244	2 F69260	nitrate ABC transp
38	31	77.5	249	2 T35589	probable secreted
39	31	77.5	269	2 T38931	hypothetical prote
40	31	77.5	303	2 C84914	hypothetical prote
41	31	77.5	372	2 A98157	probable permease
42	31	77.5	372	2 AH3130	ABC transporter, m
43	31	77.5	392	2 D83934	hypothetical prote
44	31	77.5	447	2 C84306	hypothetical prote
45	31	77.5	493	2 A71875	hypothetical prote

ALIGNMENTS

RESULT 1

A53968
serine proteinase SCCE precursor - human
N:Alternate names: stratum corneum chymotryptic enzyme
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
C:Accession: A53968
R:Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, J. Biol. Chem. 269, 19420-19426, 1994
A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic en
A:Reference number: A53968; MUID:94308225
A:Accession: A53968
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253 <HAN>
A:Cross-references: GB:L33404; NID:g521214; PIDN:AAC37551.1; PID:g532504
C:Genetics:
A:Gene: GDB:PRSS6; SCCE
A:Cross-references: GDB:377730
A:Map position: 7q35-7q35
C:Superfamily: trypsin; trypsin homology
F:30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
| | | | | | | | | |
Db 5 LLLPLQILL 13

RESULT 2

H75201
hypothetical protein PAB0088 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 08-Sep-2000
C:Accession: H75201
R:anonymous, Genoscope
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome
A:Reference number: A75001
A:Accession: H75201
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49063.1; PID:e151
A:Experimental source: strain Orsay
C:Genetics:
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0088

Query Match	90.08;	Score 36;	DB 2;	Length 146;	
Best Local Similarity	77.88;	Pred. No. 6.5;			
Matches	7;	Conservative	2;	Mismatches	0;
	Indels	0;	Gaps	0;	
Qy	1	LLLLPLQILL 9			
		:::			
Db	67	LLLLPLQIII 75			
		:::			
RESULT 3					
H84314					
Cytochrome aa3 controlling protein [imported] - Halobacterium sp. NRC-1					
C:Species: Halobacterium sp. NRC-1					
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001					
C:Accession: H84314					
R:Ng, W.V.; Kennedy, S.P.; Mahaitas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon					
Jung, K.H.; Alam, M.; Freitas, T.					
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000					
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li					
A:Title: Genome sequence of Halobacterium species NRC-1.					
A:Reference number: A84160; MUID:20504483					
A:Accession: H84314					
A>Status: preliminary					
A:Molecule type: DNA					
A:Residues: 1-277 <SPTO>					
A:Cross-references: GB:AE004437; NID:g10581096; PIDN:AAG19884.1; GSPDB:GN00138					
C:Genetics:					
A:Gene: ccp					
Query Match	85.04;	Score 34;	DB 2;	Length 277;	
Best Local Similarity	66.74;	Pred. No. 31;			
Matches	6;	Conservative	3;	Mismatches	0;
	Indels	0;	Gaps	0;	
Qy	1	LLLLPLQILL 9			
		: :::			
Db	97	VLLPLQVIL 105			
		:::			
RESULT 4					
S33603					
surfactant protein D - bovine					
C:Species: Bos primigenius taurus (cattle)					
C>Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999					
C:Accession: S33603					
R:Lim, B.L.; Lu, J.; Reid, K.B.M.					
Immunology 78, 159-165, 1993					
A:Title: Structural similarity between bovine conglutinin and bovine lung surfactant pro					
A:Reference number: S33603; MUID:93170856					
A:Accession: S33603					
A>Status: preliminary					
A:Molecule type: mRNA					
A:Residues: 1-369 <LIM>					
C:Superfamily: pulmonary surfactant protein D; C-type lectin homology					
F:248-367/Domain: C-type lectin homology <LCH>					
Query Match	85.08;	Score 34;	DB 2;	Length 369;	
Best Local Similarity	77.88;	Pred. No. 42;			
Matches	7;	Conservative	1;	Mismatches	0;
	Indels	0;	Gaps	0;	
Qy	1	LLLLPLQILL 9			
		:::			
Db	2	LLLLPLSVLL 10			
		:::			
RESULT 5					
JN0450					
conglutinin precursor - bovine					
N:Alternate names: C3b-binding protein					
N:Contains: conglutinin-N					
C:Species: Bos primigenius taurus (cattle)					
C>Date: 10-Sep-1995 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000					
C:Accession: JN0450; JC2396; S33235; A23740; S36879; S35044; I46010; A29416; S34054					

C:Superfamily: pulmonary surfactant protein D; C-type lectin homology
C:Keywords: calcium binding; glycoprotein; hydroxylysine; hydroxyproline
F:1-20/Domain: signal sequence #status predicted <SIG>
F:21-371/Product: conglutinin #status predicted <WAT>
F:46-214/Region: collagen-like

F:75-371/Product: conglutinin-N #status predicted <MA2>

F:248-369/Domain: C-type lectin homology <LCH>

F:63,87,99,135,141,159,162,198,210/Binding site: carbohydrate (Lys) (covalent) #status predicted <LCH>
F:63,87,99,135,141,159,162,198,210/Modified site: 5-hydroxylysine (Lys) #status experimental
F:78,96,108,111,129,132,147,153,171,195/Modified site: 4-hydroxyproline (Pro) #status experimental
F:337/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 85.0%; Score 34; DB 1; Length 371;

Best Local Similarity 77.8%; Pred. No. 42;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||:||

Db 2 LLLPLSVLL 10

RESULT 6

I45878

conglutinin - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 20-Aug-1999

C:Accession: I45878

R:Liu, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry, Gene 141, 277-281, 1994

A:Title: Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterization of

A:Reference number: I45878; MUID:94215917

A:Accession: I45878

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-371 <LIO>

A:Cross-references: GB:I45878; NID:g495012; PIDN:AAA20126.1; PID:g495013

C:Superfamily: pulmonary surfactant protein D; C-type lectin homology

F:248-369/Domain: C-type lectin homology <LCH>

Query Match 85.0%; Score 34; DB 2; Length 371;

Best Local Similarity 77.8%; Pred. No. 42;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||:||

Db 2 LLLPLSVLL 10

RESULT 7

AE0614

probable competence-related protein STY0984 [imported] - Salmonella enterica subsp. enter

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: this species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001

C:Accession: AE0614

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher, th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar, S.; Moulle, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov

A:Reference number: AB0502; PMID:11677608

A:Accession: AE0614

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-754 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD053383.1; PID:g16502146; GSPDB:GN00176

C:Genetics:

A:Gene: STY0984

Query Match

Best Local Similarity 85.0%; Score 34; DB 2; Length 754;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||:||

Db 371 LLLPLQVAL 379

RESULT 8

D97163

cation transport P-type ATPase CAC2137 [imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: D97163

R:Moelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;

J. Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: D97163

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-845 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80095.1; PID:g15025128; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2137

Query Match

Best Local Similarity 85.0%; Score 34; DB 2; Length 845;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
|||||:||

Db 670 LLLPLQILL 677

RESULT 9

G65039

hypothetical protein b2612 - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999

C:Accession: G65039

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.

.A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G65039

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-196 <BLAT>

A:Cross-references: GB:AE000347; GB:U00096; NID:g2367142; PIDN:AACT5661.1; PID:g1788

A:Experimental source: strain K-12, substrain MG1655

Query Match

Best Local Similarity 82.5%; Score 33; DB 2; Length 196;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||:||

Db 123 LLLPLQILL 131

RESULT 10

E70890

hypothetical protein Rv1978 - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: E70890

R:Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
 A:Reference number: A70500; MUID:98295987

A:Accession: E70890
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-282 <COL>

A:Cross-references: GB:AL022073; GB:AL123456; NID:g3256024; PIDN:CAA17851.1; PID:e125651
 A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV1978

Query Match 82.5%; Score 33; DB 2; Length 282;

Best Local Similarity 77.8%; Pred. No. 50;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||:||

Db 196 LLLPLHLLL 204

RESULT 11

AB3334 daunorubicin resistance transmembrane protein [imported] - Brucella melitensis (strain 1

C:Species: Brucella melitensis

C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002

R:DelVecchio, V.G.; Kapral, R.J.; Redkar, R.J.; Patra, G.; Mujeer, C.; Los, T.; Ivanova,

.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess

proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis

A:Reference number: AD3252; PMID:11756688

A:Accession: AB3334

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-370 <KUR>

A:Cross-references: GB:AE008917; PIDN:AAL51837.1; PID:g17982584; GSPDB:GN00190

A:Experimental source: strain 16M

C:Genetics:

A:Gene: BMEI0656

A:Map position: I

Query Match 82.5%; Score 33; DB 2; Length 370;

Best Local Similarity 75.0%; Pred. No. 66;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQIL 8

|||||:|

Db 293 ILLPLQVL 300

RESULT 12

C91063 hypothetical protein EGS3475 [imported] - Escherichia coli (strain O157:H7, substrain RI

C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 24-Aug-2001

C:Accession: C91063

R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc

A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: C91063

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-398 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA36898.1; PID:g13362946; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: EGS3475

C:Superfamily: hypothetical protein HI0107

Query Match 82.5%; Score 33; DB 2; Length 398;

Best Local Similarity 77.8%; Pred. No. 71;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQIL 8

|||||:||

Db 301 LLLPAQLLL 309

RESULT 15

A30083 hypothetical protein orf470 [imported] - Giardia theta nucleomorph

C:Species: nucleomorph Giardia theta

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||:|

Db 101 LLLPLQILL 109

RESULT 13

AC0834 probable membrane protein corB [imported] - Salmonella enterica subsp. enterica serov

C:Species: Salmonella enterica subsp. enterica serovar Typhi

A:Note: This species has also been called Salmonella typhi

C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001

C:Accession: AC0834

R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church

th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr

, S.; Moulder, S.; O'Gaora, P.

Nature 413, 848-852, 2001

A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,

A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica se

A:Reference number: AB0502; PMID:11677608

A:Accession: AC0834

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-413 <PAR>

A:Cross-references: GB:AL513382; PIDN:CAD05858.1; PID:g16503833; GSPDB:GN00176

C:Genetics:

A:Gene: corB

C:Superfamily: hypothetical protein HI0107

Query Match 82.5%; Score 33; DB 2; Length 413;

Best Local Similarity 77.8%; Pred. No. 74;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||:|

Db 116 LLLPLQILL 124

RESULT 14

C83103 hypothetical protein PA4338 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: C83103

R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.

adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.

.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p

A:Reference number: AB2950; MUID:20437337

A:Accession: C83103

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-426 <STO>

A:Cross-references: GB:AE004850; GB:AE004091; NID:g9950560; PIDN:AAG07726.1; GSPDB:C

A:Experimental source: strain PA01

C:Genetics:

A:Gene: PA4338

Query Match 82.5%; Score 33; DB 2; Length 426;

Best Local Similarity 77.8%; Pred. No. 76;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||:||

Db 301 LLLPAQLLL 309

RESULT 15

A30083 hypothetical protein orf470 [imported] - Giardia theta nucleomorph

C:Species: nucleomorph Giardia theta

A: Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
 C: Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
 C: Accession: A90083
 R: Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Rea
 Nature 410, 1091-1096, 2001
 A: Title: The highly reduced genome of an enslaved algal nucleus.
 A: Reference number: A9082; MUID:11323671
 A: Accession: A90083
 A: Status: preliminary
 A: Molecule type: DNA
 A: Residues: 1-470 <DOU>
 A: Cross-references: GB:AF165818; NID:g6690603; PIDN:AAF24211.1; GSPDB:GN00150
 C: Genetics:
 A: Gene: orf470
 A: Map position: 1
 A: Genome: nucleomorph
 C: Keywords: nucleomorph

Query Match 82.5%; Score 33; DB 2; Length 470;
 Best Local Similarity 66.7%; Pred. No. 84;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 :||||:|
 DB 253 MLLPLEILI 261

RESULT 16
 A69149
 O-antigen transporter related protein - Methanobacterium thermoautotrophicum (strain Del
 C: Species: Methanobacterium thermoautotrophicum
 C: Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
 C: Accession: A69149
 R: Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.;
 ; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiواني, N.
 ki, S.; Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7155, 1997
 A: Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: func
 A: Reference number: A69000; MUID:98037514
 A: Accession: A69149
 A: Status: preliminary; nucleic acid sequence not shown; translation not shown
 A: Molecule type: DNA
 A: Residues: 1-475 <MTH>
 A: Cross-references: GB:AE000823; GB:AE000666; NID:g2621432; PIDN:AAB84885.1; PID:g262144
 A: Experimental source: strain Delta H
 C: Genetics:
 A: Gene: MTH379
 A: Start codon: GTG
 C: Superfamily: succinoglycan biosynthesis transport.protein

Query Match 82.5%; Score 33; DB 1; Length 475;
 Best Local Similarity 77.8%; Pred. No. 85;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 :||||:|
 DB 432 LLLPLSLILI 440

RESULT 17
 A46362
 amyloid precursor-like protein - mouse
 C: Species: Mus musculus (house mouse)
 C: Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
 C: Accession: A46362
 R: Wasco, W.; Bupp, K.; Magendantz, M.; Gusella, J.F.; Tanzi, R.E.; Solomon, F.
 Proc. Natl. Acad. Sci. U.S.A. 89, 10758-10762, 1992
 A: Title: Identification of a mouse brain cDNA that encodes a protein related to the Alzh
 A: Reference number: A46362; MUID:93066322
 A: Accession: A46362
 A: Status: preliminary
 A: Molecule type: nucleic acid
 A: Residues: 1-653 <WAS>

A: Experimental source: brain
 A: Note: sequence inconsistent with the nucleotide translation
 A: Note: sequence extracted from NCBI backbone (NCBIN:118683, NCBIP:118684)
 C: Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type proteina
 C: Keywords: transmembrane protein

Query Match 82.5%; Score 33; DB 2; Length 653;
 Best Local Similarity 77.8%; Pred. No. 1.2e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 :||||:|
 DB 21 LLLPLSLLL 29

RESULT 18
 S25656
 T-cell surface glycoprotein CD8 alpha chain precursor - orangutan
 C: Species: Pongo pygmaeus (orangutan)
 C: Date: 26-Feb-1998 #sequence_revision 26-Feb-1998 #text_change 23-Jul-1999
 C: Accession: S25656
 R: Lawlor, D.A.; Parham, P.
 Immunogenetics 36, 121-125, 1992
 A: Title: Structure of CD8 alpha and beta chains of the orangutan: novel patterns of
 A: Reference number: S25656; MUID:92307742
 A: Accession: S25656
 A: Molecule type: mRNA
 A: Residues: 1-198 <LAW>
 A: Cross-references: EMBL:X60223; NID:g38144; PIDN:CAA42784.1; PID:g38145
 C: Superfamily: immunoglobulin V region; immunoglobulin homology
 C: Keywords: alternative splicing; extracellular protein; glycoprotein; heterotetramer
 F: 1-21/Domain: signal sequence #status predicted <SIG>
 F: 22-198/Product: T-cell surface glycoprotein CD8 alpha chain #status predicted <MAT
 F: 36-117/Domain: immunoglobulin homology <IMM>
 F: 146-168/Domain: transmembrane #status predicted <TM>
 F: 169-198/Domain: intracellular #status predicted <INT>
 F: 43-115/Disulfide bonds: #status predicted
 F: 49/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F: 144/Disulfide bonds: interchain #status predicted

Query Match 80.0%; Score 32; DB 2; Length 198;
 Best Local Similarity 77.8%; Pred. No. 55;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 :||||:|
 DB 8 LLLPLALLL 16

RESULT 19
 RWUT8
 T-cell surface glycoprotein CD8 alpha chain precursor - human
 N: Alternate names: Leu-2/T8 T lymphocyte differentiation antigen; T-cell surface anti
 N: Contains: T-cell surface glycoprotein CD8 alpha chain secreted splice form; T-cell
 C: Species: Homo sapiens (man)
 C: Date: 28-May-1986 #sequence_revision 28-May-1986 #text_change 22-Jun-1999
 C: Accession: A30604; A45888; A01999; A22824; A50096; A31458; B31458; JP0105
 R: Normont, A.M.; Lonberg, N.; Lacy, E.; Littman, D.R.
 J. Immunol. 142, 3312-3319, 1989
 A: Title: Alternatively spliced mRNA encodes a secreted form of human CD8-alpha. Char.
 A: Reference number: A30604; MUID:89215302
 A: Accession: A30604
 A: Molecule type: DNA
 A: Residues: 1-235 <NOR>
 A: Cross-references: GB:M26315; NID:g341466; PIDN:AAA79217.1; PID:g1019167
 R: Nakayama, K.; Tokito, S.; Okumura, K.; Nakauchi, H.
 Immunogenetics 30, 393-397, 1989
 A: Title: Structure and expression of the gene encoding CD8-alpha chain (Leu-2/T8).
 A: Reference number: A45888; MUID:90035142
 A: Accession: A45888
 A: Molecule type: DNA
 A: Residues: 1-235 <NAK>
 A: Cross-references: GB:M27161; NID:g187844; PIDN:AAA59674.1; PID:g386908

A:Note: the authors translated the codon TCG for residue 116 as Cys and GCC for residue 117 as Arg.
 R:Littman, D.R.; Thomas, Y.; Maddon, P.J.; Chess, L.; Axel, R.
 Cell 40, 237-246, 1985
 A:Title: The isolation and sequence of the gene encoding T8: a molecule defining function for T8 lymphocytes.
 A:Reference number: A01999; MUID:85099337
 A:Accession: A01999
 A:Molecule type: mRNA
 A:Residues: 1-235 <LIT>
 A:Cross-references: GB:M12828; NID:gl79145; PIDN:AAB04637.1; PID:gl79146
 A:Experimental source: clones pT8.B and pT8.F1
 R:Sukhatme, V.P.; Sizer, K.C.; Vollmer, A.C.; Hunkapiller, T.; Parnes, J.R.
 Cell 40, 591-597, 1985
 A:Title: The T cell differentiation antigen Leu-2/T8 is homologous to immunoglobulin and has a unique structure.
 A:Reference number: A22824; MUID:85124610
 A:Accession: A22824
 A:Molecule type: mRNA
 A:Residues: 1-235 <SUK>
 A:Cross-references: GB:M12824; NID:g339426; PIDN:AAA61133.1; PID:g339427
 R:Parnes, J.R.; Sizer, K.C.; Sukhatme, V.P.; Hunkapiller, T.
 Behring Inst. Mitt. 77, 48-55, 1985
 A:Title: Structure of Leu-2/T8 as deduced from the sequence of a cDNA clone.
 A:Reference number: A90096; MUID:86103103
 A:Accession: A90096
 A:Molecule type: mRNA
 A:Residues: 1-235 <PAR>
 R:Giblin, P.; Ledbetter, J.A.; Kavathas, P.
 Proc. Natl. Acad. Sci. U.S.A. 86, 998-1002, 1989
 A:Title: A secreted form of the human lymphocyte cell surface molecule CD8 arises from a single gene.
 A:Reference number: A31458; MUID:89128905
 A:Accession: A31458
 A:Molecule type: mRNA
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Residues: 168-235 <GTB>
 A:Note: this mRNA fragment represents the transmembrane form
 A:Accession: B31458
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Residues: 168-171, 'G', 210-235 <GI2>
 A:Note: this mRNA fragment represents the secreted form
 A:Note: the cited Genbank accession number, J04165, is not in release
 R:Snow, P.M.; Terhorst, C.
 J. Biol. Chem. 258, 14675-14681, 1983
 A:Reference number: A92407; MUID:84061928
 A:Contents: annotation
 C:Comment: This protein was shown in reference A92407 to be in a multimeric complex of the T cell surface molecule CD8.
 C:Genetics:
 A:Gene: GDB:CD8A
 A:Cross-references: GDB:120581; OMIM:186910
 A:Map position: 2p12-2p13
 A:Introns: 17/1; 135/1; 172/1; 209/1; 219/2
 C:Superfamily: immunoglobulin v region; immunoglobulin homology
 C:Keywords: alternative splicing; extracellular protein; glycoprotein; heterotetramer; 1
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-235/Product: T-cell surface glycoprotein CD8 alpha chain transmembrane splice form
 F:22-171, 209-235/Product: T-cell surface glycoprotein CD8 alpha chain secreted splice form
 F:36-117/Domain: immunoglobulin homology <IMV>
 F:118-182/Domain: hinge <HNG>
 F:183-205/Domain: transmembrane #status predicted <TMV>
 F:206-235/Domain: intracellular #status predicted <INT>
 F:43-115/Disulfide bonds: #status predicted
 F:49/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:181/Disulfide bonds: interchain #status predicted

Query Match 80.0%; Score 32; DB 1; Length 235;
 Best Local Similarity 77.8%; Pred. No. 65;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
 ||||| :||
 Db 8 LLLPLALLL 16

RESULT 20

LNRTMA

N:Alternate names: serum mannan-binding protein
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 22-Jun-1999
 C:Accession: B24791; A29556; A27799
 R:Drickamer, K.; Dordal, M.S.; Reynolds, L.
 J. Biol. Chem. 261, 6878-6887, 1986
 A:Title: Mannose-binding proteins isolated from rat liver contain carbohydrate-recognizing sites.
 A:Reference number: A24791; MUID:86196130
 A:Accession: B24791
 A:Molecule type: mRNA
 A:Residues: 1-238 <DR1>
 R:Drickamer, K.; McCreary, V.
 J. Biol. Chem. 262, 2582-2589, 1987
 A:Title: Exon structure of a mannose-binding protein gene reflects its evolutionary r
 A:Reference number: A29556; MUID:87137502
 A:Accession: A29556
 A:Molecule type: DNA
 A:Residues: 1-155, 'K', 157-238 <DR2>
 A:Cross-references: GB:M14104; NID:g205259; PIDN:AAA98781.1; PID:g205261
 A:Note: the codon AAG for residue 156 is inconsistent with the authors' statement tha
 R:Ikeda, K.; Sannoh, T.; Kawasaki, N.; Kawasaki, T.; Yamashina, I.
 J. Biol. Chem. 262, 7451-7454, 1987
 A:Title: Serum lectin with known structure activates complement through the classical
 A:Reference number: A27799; MUID:87222358
 A:Accession: A27799
 A:Molecule type: protein
 A:Residues: 18-42 <IKE>
 C:Comment: Mannose-binding lectins are opsonins that are important in host defense ag
 C:Comment: This plasma protein binds mannose and N-acetylglucosamine and can activate
 C:Comment: The molecule consists of approximately 20 identical chains linked by disul
 C:Genetics:
 A:Introns: 51/1; 90/1; 115/1
 C:Superfamily: mannose-binding lectin; C-type lectin homology
 C:Keywords: acute phase; calcium; hydroxyllysine; hydroxyproline; lectin; liver; plasm
 F:1-17/Domain: signal sequence #status predicted <SIG>
 F:18-238/Product: mannose-binding lectin A #status experimental <MAT>
 F:36-88/Region: collagen-like
 F:85-87/Region: cell attachment (R-G-D) motif
 F:118-234/Domain: C-type lectin homology <LCH>
 F:61, 67, 73/Modified site: 4-hydroxyproline (Pro) #status experimental
 F:79, 82/Modified site: lysine derivative (Lys) (probably 5-hydroxylysine) #status exp
 Query Match 80.0%; Score 32; DB 1; Length 238;
 Best Local Similarity 77.8%; Pred. No. 66;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY

Db

RESULT 21

KYHUCM

chymase (EC 3.4.21.39) precursor [validated] - human
 N:Alternate names: chymase I; mast cell proteinase 1
 C:Species: Homo sapiens (man)
 C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 08-Dec-2000
 C:Accession: A39861; A39861; S15906; S33247; S56063; A23686; A60772
 R:Urata, H.; Kinoshita, A.; Perez, D.M.; Misono, K.S.; Bumpus, F.M.; Graham, R.M.; Hu
 J. Biol. Chem. 266, 17173-17179, 1991
 A:Title: Cloning of the gene and cDNA for human heart chymase.
 A:Reference number: A40967; MUID:91373329
 A:Accession: A40967
 A:Molecule type: mRNA
 A:Residues: 1-247 <URA>
 A:Cross-references: GB:M69136; NID:gl80539; PIDN:AAA52019.1; PID:gl80540
 A:Experimental source: ventricular myocardium
 A:Accession: B40967
 A:Molecule type: DNA

A:Residues: 1-247 <UR2>
A:Cross-references: GB:M69137; NID:gl80543; PIDN:AAA52021.1; PID:gl80544
A:Note: part of this sequence, including the amino end of the mature protein, was confirmed by R. Caughey, G.H.; Zerweck, E.H.; Vanderslice, P.
R:Biochem. J. 266, 12956-12963, 1991
A:Title: Structure, chromosomal assignment, and deduced amino acid sequence of a human 9
A:Reference number: A39861; MUID:91302311
A:Accession: A39861
A:Molecule type: DNA
A:Residues: 1-247 <CAU>
A:Cross-references: GB:M64269; NID:gl80541; PIDN:AAA52020.1; PID:gl80542
A:Note: these authors suggest that mast cells are the source of heart chymase
R:Jenne, D.E.; Tschopp, J.
Biochem. J. 276, 567-568, 1991
A:Title: Angiotensin II-forming heart chymase is a mast-cell-specific enzyme.
A:Reference number: S15906; MUID:91264818
A:Accession: S15906
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 26-60 <JEN>
A:Cross-references: EMBL:X59072; NID:g29949; PIDN:CAA41796.1; PID:g29950
R:Sukenaga, Y.; Kido, H.; Neki, A.; Enomoto, M.; Ishida, K.; Takagi, K.; Katunuma, N.
FEBS Lett. 323, 119-122, 1993
A:Title: Purification and molecular cloning of chymase from human tonsils.
A:Reference number: S33247; MUID:93265916
A:Accession: S33247
A:Molecule type: mRNA
A:Residues: 22-27, 'S', 29-247 <SUK>
A:Cross-references: GB:S61334; NID:g409010; PIDN:AAB26828.1; PID:g409011
A:Accession: S56063
A:Molecule type: protein
A:Residues: 22-27, 'S', 29-42 <SU2>
A:Experimental source: tonsil
R:Draca, H.; Kinoshita, A.; Misono, K.S.; Bumpus, F.M.; Husain, A.
J. Biol. Chem. 265, 22348-22357, 1990
A:Title: Identification of a highly specific chymase as the major angiotensin II-forming
A:Reference number: A23686; MUID:91093078
A:Accession: A23686
A:Molecule type: protein
A:Residues: 22-30, 32-39, 'Q', 41-45, 136-141, 'I', 143 <UR3>
R:Schwartz, N.M.; Irani, A.M.A.; Sprows, J.L.; Abernethy, J.; Wintroub, B.; Schwartz, I.
J. Immunol. 145, 2652-2661, 1990
A:Title: Identification of a cathepsin G-like proteinase in the MC-TC type of human mast
A:Reference number: A60772; MUID:91010743
A:Accession: A60772
A:Molecule type: protein
A:Residues: 22-27, 'S', 29-56 <SCH>
C:Comment: Functions of this chymotrypsin-like serine proteinase after mast cell degradation
he cleavage of angiotensin I to angiotensin II is accomplished by this enzyme rather than
C:Genetics:
A:Gene: GDB:CMAL
A:Cross-references: GDB:127603; OMIM:118940
A:Map position: 14q11.2-14q11.2
A:Introns: 20/1; 70/2; 115/3; 200/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-21/Domain: propeptide #status predicted <PRO>
F:22-247/Product: chymase #status experimental <MAT>
F:22-240/Domain: trypsin homology <TRY>
F:66,110,203/Active site: His, Asp, Ser #status predicted
F:80/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:103/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 80.0%; Score 32; DB 1; Length 247;
Best Local Similarity 77.8%; Pred. No. 69;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLILL 9
|||||: ||
Db 2 LLLPLPLLL 10

RESULT 22
F91095
type III secretion protein EivF [imported] - Escherichia coli (strain O157:H7, substr
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: F91095
R:Hayashi, T.; Makino, K.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.; Han, C.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F91095
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA837157.1; PID:gl3363206; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 050952
C:Genetics:
A:Gene: ECs3734

Query Match 80.0%; Score 32; DB 2; Length 249;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLILL 9
|||||: ||
Db 6 LLLPMNVLL 14

RESULT 23
B85941
hypothetical protein Z4198 [imported] - Escherichia coli (strain O157:H7, substrain F
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: B85941
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May
riller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A89480; MUID:21074935; PMID:11206551
A:Accession: B85941
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-249 <STO>
A:Cross-references: GB:AE005174; NID:g12517376; PIDN:AAG57990.1; GSPDB:GN00145; UWG
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z4198

Query Match 80.0%; Score 32; DB 2; Length 249;
Best Local Similarity 66.7%; Pred. No. 69;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLILL 9
|||||: ||
Db 6 LLLPMNVLL 14

RESULT 24
C97402
probable acyltransferase (AF232919) [imported] - Agrobacterium tumefaciens (strain C5
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: C97402
R:Goodner, B.; Hinkley, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldr
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: C97402
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-264 <KUR>

A:Cross-references: GB:AE007869; PIDN:AAK86172.1; PID:g15155265; GSPDB:GN00169

C:Genetics:

A:Gene: AGR_C_621

A:Map position: circular chromosome

Query Match 80.0%; Score 32; DB 2; Length 264;

Best Local Similarity 75.0%; Pred. No. 73;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQL 8

Db 19 ILLPLQLL 26

RESULT 25

AC2620

1-acyl-sn-glycerol-3-phosphate acyltransferase plsc [imported] - Agrobacterium tumefaciens

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AC2620

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, H.

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AC2620

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-264 <KUR>

A:Cross-references: GB:AE008688; PIDN:AA141377.1; PID:g17738693; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: plsc

A:Map position: circular chromosome

Query Match 80.0%; Score 32; DB 2; Length 264;

Best Local Similarity 75.0%; Pred. No. 73;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQL 8

Db 19 ILLPLQLL 26

RESULT 26

JC7300

tax-responsive element-binding protein 5 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000

C:Accession: JC7300

R:Masaki, T.; Noguchi, H.; Kobayashi, M.; Yoshida, M.; Takamatsu, K.

DNA Res. 7, 187-193, 2000

A:Title: Isolation and characterization of the gene encoding mouse tax-responsive element

A:Reference number: JC7300

A:Accession: JC7300

A:Molecule type: mRNA

A:Residues: 1-266 <MAS>

A:Cross-references: DDBJ:AB036745

C:Genetics:

A:Gene: treb5

C:Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C:Keywords: leucine zipper; transcription factor

Query Match

Best Local Similarity 80.0%; Score 32; DB 2; Length 266;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 8

Db 187 LLLPLQLL 193

RESULT 27

JC4857

hepatocarcinogenesis-related transcription factor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 08-Sep-1997

C:Accession: JC4857

R:Kishimoto, T.; Kokura, K.; Kumagai, Y.; Makino, Y.; Tamura, T.

Biochem. Biophys. Res. Commun. 224, 746-751, 1996

A:Title: HTF: A b-zip transcription factor that is closely related to the human XBP/T

A:Reference number: JC4857

A:Accession: JC4857

A:Molecule type: mRNA

A:Residues: 1-267 <KIS>

C:Comment: This is a basic-leucine zipper type transcription factor involved in hepat

C:Genetics:

A:Gene: htf

C:Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C:Keywords: leucine zipper; transcription factor

F:58-98/Domain: fos/jun DNA-binding domain homology <FJD>

F:89-126/Region: leucine zipper motif

Query Match

Best Local Similarity 80.0%; Score 32; DB 2; Length 267;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 8

Db 188 LLLPLQLL 194

RESULT 28

JC6197

stromelysin 3 (EC 3.4.24.-) - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 17-Mar-1999

C:Accession: JC6197

R:Okada, A.; Saez, S.; Misumi, Y.; Basset, P.

Gene 185, 187-193, 1997

A:Title: Rat stromelysin 3: cDNA cloning from healing skin wound, activation by furin

A:Reference number: JC6197; MUID:97208872

A:Contents: Skin wounds

A:Accession: JC6197

A:Molecule type: mRNA

A:Residues: 1-491 <OKA>

A:Cross-references: GB:U46034

C:Comment: This protein is a member of the matrix metalloproteinase family.

C:Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metallopro

C:Keywords: hydrolase; metalloproteinase; zinc; zymogen

F:52-261/Domain: matrix metalloproteinase homology <MMP>

F:291-483/Domain: hemopexin repeat homology <PXN>

F:84-218,222,228/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #sta

F:218,222,228/Binding site: zinc, catalytic (His) (active) #status predicted

F:219/Active site: Glu #status predicted

Query Match

Best Local Similarity 80.0%; Score 32; DB 2; Length 491;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9

Db 15 LLLPLPLLL 23

RESULT 29

A44399

stromelysin 3 (EC 3.4.24.-) - mouse

N:Alternate names: matrix metalloproteinase 11 (MMP11)

C:Species: Mus musculus (house mouse)

C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 10-Jul-1998

C:Accession: A44399

R:Lefebvre, O.; Wolf, C.; Llimacher, J.M.; Hutin, P.; Wendling, C.; LeMeur, M.; Basset

J. Cell Biol. 119, 997-1002, 1992
A:Title: The breast cancer-associated stromelysin-3 gene is expressed during mouse mammary gland development
A:Reference number: A44399; MUID:93054930
A:Accession: A44399
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-492 <LEF>
A:Note: sequence extracted from NCBI backbone (NCBIP:117216)
C:Superfamily: Interstitial collagenase; hemopexin repeat homology; matrix metalloproteinase
C:Keywords: hydrolase; metalloproteinase; zinc; zymogen
F:52-262/Domain: matrix metalloproteinase homology <MMP>
F:292-484/Domain: hemopexin repeat homology <PN>
F:84,219,223,229/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status
F:219,223,229/Binding site: zinc, catalytic (His) (active) #status predicted
F:220/Active site: Glu #status predicted

Query Match 80.0%; Score 32; DB 2; Length 492;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
||||| :||
Db 15 LLLPLPLLL 23

RESULT 30
A:Title: Na+/H+-exchanging protein [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: A12088
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. PCC 7120
A:Reference number: A1807; MUID:21595285; PMID:11759840
A:Accession: A12088
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-543 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA073963.1; PID:g17131355; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2264

Query Match 80.0%; Score 32; DB 2; Length 543;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
||||| :||
Db 194 LLLPLQILL 201

RESULT 31
A:Title: peptidyl-dipeptidase A (EC 3.4.15.1) precursor, renal and pulmonary splice form - human
N:Alternate names: angiotensin I-converting enzyme (ACE); CD143; dipeptidyl carboxypeptidase
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1990 #sequence_revision 02-Jul-1998 #text_change 18-Jun-1999
C:Accession: A31759; P00004
R:Soubrier, F.; Alhenc-Gelas, F.; Hubert, C.; Allegrini, J.; John, M.; Tregear, G.; Corvol, A.C.
Proc. Natl. Acad. Sci. U.S.A. 85, 9386-9390, 1988
A:Title: Two putative active centers in human angiotensin I-converting enzyme revealed by X-ray crystallography
A:Reference number: A31759; MUID:89071703
A:Accession: A31759
A:Molecule type: mRNA
A:Residues: 1-1306 <SOU>
A:Cross-references: GB:J04144; NID:g178285; PIDN:AAA51684.1; PID:g178286
A:Experimental source: kidney
A:Note: parts of this sequence, including the amino end of the mature protein, were determined from a cDNA library
R:Takeuchi, K.; Shimizu, T.; Ohishi, N.; Seyama, Y.; Takaku, F.; Yotsumoto, H.

J. Biochem. 106, 442-445, 1989
A:Title: Purification of human lung angiotensin-converting enzyme by high-performance liquid chromatography
A:Reference number: P00004; MUID:90110025
A:Accession: P00004
A:Molecule type: protein
A:Residues: 'XX',32-34,'E',36-37,'X',39-41,'R',43-46 <TAK>
A:Experimental source: lung
C:Comment: This splice form is found in many tissues, in particular kidney and lung
C:Genetics:
A:Gene: GDB:DCPI; ACE
A:Cross-references: GDB:119840; OMIM:106180
A:Map position: 17q23-17q23
C:Function:
A:Note: plays a role in the control of blood pressure by catalyzing the conversion of angiotensin I to angiotensin II
C:Superfamily: mammalian peptidyl-dipeptidase A
C:Keywords: alternative splicing; blood pressure control; glycoprotein; kidney; lung; F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-1306/Product: peptidyl dipeptidase I #status predicted <MAT>
F:1260-1276/Domain: transmembrane #status predicted <TRM>
F:38-54,74,111,146,160,318,445,509,523,677,695,714,760,942,1191,1225/Binding site: Cys
F:390,394/Binding site: zinc (His) #status predicted
F:988,992,1008/Binding site: zinc, catalytic (His, His, Glu) #status predicted
F:989/Active site: Glu #status predicted

Query Match 80.0%; Score 32; DB 1; Length 1306;
Best Local Similarity 77.8%; Pred. No. 3.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
||||| :||
Db 12 LLLPLPLLL 20

RESULT 32
A:Title: apolipoprotein C-II precursor - guinea pig
C:Species: Cavia porcellus (guinea pig)
C:Date: 06-Mar-1992 #sequence_revision 10-Apr-1992 #text_change 13-Aug-1999
C:Accession: A38685
R:Andersson, Y.; Thelander, L.; Bengtsson-Olivercrona, G.
J. Biol. Chem. 266, 4074-4080, 1991
A:Title: Demonstration of apolipoprotein CII in guinea pigs. Functional characterization of the CII gene
A:Reference number: A38685; MUID:91154195
A:Accession: A38685
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-100 <AND>
A:Cross-references: GB:M59913; NID:g191238; PIDN:AAA37031.1; PID:g191239
A:Note: the authors translated the codon GAG for residue 68 as Gln
C:Superfamily: apolipoprotein A-I
C:Keywords: lipid binding; lipoprotein

Query Match 77.5%; Score 31; DB 2; Length 100;
Best Local Similarity 77.8%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
||||| :||
Db 11 LLLPLPLLL 19

RESULT 33
A:Title: interleukin-2 precursor - rat
N:Alternate names: IL-2; T-cell growth factor
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-Apr-1989 #sequence_revision 26-Apr-1989 #text_change 16-Jul-1999
C:Accession: A45882; A31278
R:McKnight, A.J.; Mason, D.W.; Barclay, A.N.
Immunogenetics 30, 145-147, 1989
A:Title: Sequence of rat interleukin 2 and anomalous binding of a mouse interleukin 2 receptor to rat interleukin 2
A:Reference number: A45882; MUID:89339608

A:Accession: A45882

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-155 <WCK>

A:Cross-references: GB:M22899; NID:g204909; PIDN:AAA41427.1; PID:g204910

C:Superfamily: Interleukin-2

C:Keywords: cytokine; glycoprotein; growth factor; immunoregulation; T-cell

Query Match 77.5%; Score 31; DB 2; Length 155;

Best Local Similarity 77.8%; Pred. No. 67;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

||||| :||

Db 37 LLLDLQVLL 45

RESULT 34

G75555

conserved hypothetical protein - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000

C:Accession: G75555

R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;

M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma

S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.

Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.

A:Reference number: A73250; MUID:20036896

A:Accession: G75555

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-159 <WHI>

A:Cross-references: GB:AE001876; GB:AE000513; NID:g6457800; PIDN:AAF09727.1; PID:g645780

A:Experimental source: Strain R1

C:Genetics:

A:Gene: DR0138

A:Map position: 1

Query Match 77.5%; Score 31; DB 2; Length 159;

Best Local Similarity 66.7%; Pred. No. 69;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

||||| :||

Db 120 LLPLSLLL 128

RESULT 35

S10125

alpha-2u-globulin III precursor - mouse

N:Alternate names: group I major urinary protein

C:Species: Mus musculus (house mouse)

C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 20-Aug-1999

C:Accession: S10125; E26890

R:Clark, A.J.; Chave-Cox, A.; Ma, X.; Bishop, J.O.

EMBO J. 4, 3167-3171, 1985

A:Title: Analysis of mouse major urinary protein genes: variation between the exonic seq

has occurred between MUP genes.

A:Reference number: S07186; MUID:86135943

A:Accession: S10125

A>Status: translation not shown

A:Molecule type: mRNA

A:Residues: 1-184 <CLA>

A:Cross-references: EMBL:X03525; NID:g53272; PIDN:CAA27228.1; PID:g736265

A:Experimental source: clone pMUP15

R:Shahan, K.; Gilmartin, M.; Derman, E.

Mol. Cell. Biol. 7, 1938-1946, 1987

A:Title: Nucleotide sequences of liver, lachrymal, and submaxillary gland mouse major ur

bes.

A:Reference number: A93091; MUID:87257897

A:Accession: E26890

A:Molecule type: mRNA

A:Residues: 72-184 <SHA>

A:Cross-references: GB:M16359; NID:g199864; PIDN:AAA39762.1; PID:g199865

A:Experimental source: strain BALB/C, female liver

C:Superfamily: lipocalin; lipocalin homology

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-184/Product: alpha-2u-globulin #status predicted <MAT>

F:32-179/Domain: lipocalin homology <LIP>

Query Match 77.5%; Score 31; DB 2; Length 184;

Best Local Similarity 77.8%; Pred. No. 80;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

||||| :||

Db 3 LLLPLLLLL 11

RESULT 36

T43766

hypothetical protein 189 [imported] - slime mold (Dictyostellium discoideum) mitochondr

C:Species: mitochondrion Dictyostellium discoideum

C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 19-May-2000

C:Accession: T43766

R:Ogawa, S.; Yoshino, R.; Angata, K.; Pi, M.; Iwamoto, M.; Kuroe, K.; Matsuo, K.; Mor

submitted to the EMBL Data Library, December 1996

A:Description: The mitochondrial DNA of Dictyostellium discoideum. Complete sequence.

A:Reference number: Z22666

A:Accession: T43766

A>Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-189 <OGA>

A:Cross-references: EMBL:AB000109; PIDN:BAA78070.1

C:Genetics:

A:Genome: mitochondrion

C:Superfamily: Dictyostellium mitochondrion hypothetical protein 189

C:Keywords: mitochondrion

Query Match 77.5%; Score 31; DB 2; Length 189;

Best Local Similarity 75.0%; Pred. No. 82;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPLQILL 9

||||| :

Db 106 LLPLQILV 113

RESULT 37

F69260

nitrate ABC transporter, permease protein (nrtB-1) homolog - Archaeoglobus fulgidus

C:Species: Archaeoglobus fulgidus

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000

C:Accession: F69260

R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod

; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E

Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes,

Smith, H.O.; Woese, C.R.; Venter, J.C.

A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arc

A:Reference number: A69250; MUID:98049343

A:Accession: F69260

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-244 <KLE>

A:Cross-references: GB:AE001100; GB:AE000782; NID:g2689423; PIDN:AAB91143.1; PID:g26

C:Superfamily: Synchococcus nitrate transport protein nrtB

Query Match 77.5%; Score 31; DB 2; Length 244;

Best Local Similarity 77.8%; Pred. No. 11e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

||||| |||

Db 102 VLLPLYLL 110

RESULT 38

T35589

probable secreted protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 01-Dec-2000

C:Accession: T35589; S37564

R:Saunders, D.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, August 1998

A:Reference number: Z21583

A:Accession: T35589

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-249 <SAU>

A:Cross-references: EMBL:AL031317; PIDN:CRAA20416.1; GSPDB:GN00070; SCOEDB:SC6G4.38

R:Experimental source: strain A3(2)

R:Duchene, A.M.; Kieser, K.; Hopwood, D.; Thompson, C.; Mazodier, P.

submitted to the EMBL Data Library, September 1993

A:Description: Molecular characterization of two groEL genes in Streptomyces coelicolor

A:Reference number: S37564

A:Accession: S37564

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 121-249 <DUC>

A:Cross-references: EMBL:X75206; NID:g406595; PIDN:CAA53017.1; PID:g406596

C:Genetics:

A:Gene: SCOEDB:SC6G4.38

Query Match

Best Local Similarity 77.5%; Score 31; DB 2; Length 249;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||

Db 16 LLLPLLLLL 24

RESULT 39

T38931

hypothetical protein SPAC57A10.04 - fission yeast (Schizosaccharomyces pombe)

C:Species: Schizosaccharomyces pombe

C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 04-Mar-2000

C:Accession: T38931

R:Badcock, K.; Churcher, C.M.; Wood, V.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, April 1997

A:Reference number: Z21818

A:Accession: T38931

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-269 <BAD>

A:Cross-references: EMBL:Z94864; PIDN:CAB08167.1; GSPDB:GN00066; SPDB:SPAC57A10.04

A:Experimental source: strain 972h-; cosmid c57A10

C:Genetics:

A:Gene: SPDB:SPAC57A10.04

A:Map position: 1

C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC57A10.04

Query Match

Best Local Similarity 77.5%; Score 31; DB 2; Length 269;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||

Db 81 LLLPLQHL 89

RESULT 40

C84914

hypothetical protein At2g47360 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: C84914

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,

euss, D.; Niernman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487

A:Accession: C84914

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-303 <STO>

A:Cross-references: GB:AE002093; NID:g2275214; PIDN:AABG3836.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g47360

A:Map position: 2

Query Match

Best Local Similarity 77.5%; Score 31; DB 2; Length 303;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||

Db 30 LLLPLSFL 38

RESULT 41

A98157

probable permease of ABC transporter PA5230 [imported] - Agrobacterium tumefaciens (s

C:Species: Agrobacterium tumefaciens

C>Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002

C:Accession: A98157

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourollo, B.; Goldm

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz,

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium

A:Reference number: A97359; PMID:11743194

A:Accession: A98157

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-372 <KUR>

A:Cross-references: GB:AE007870; PIDN:AAK88779.1; PID:g15158527; GSPDB:GN00170

C:Genetics:

A:Gene: AGR_L_429

A:Map position: linear chromosome

Query Match

Best Local Similarity 85.7%; Score 31; DB 2; Length 372;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQIL 8

|||||

Db 296 LLLPLQL 302

RESULT 42

AH3130

ABC transporter, membrane spanning protein Atu4668 [imported] - Agrobacterium tumefac

C:Species: Agrobacterium tumefaciens

C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AH3130

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayin, T.; Levy, R.; Li, M.; MCCI

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AH3130

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-372 <KUR>

A:Cross-references: GB:AE008689; PIDN:NAL45462.1; PID:g17743167; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu4668
A:Map position: linear chromosome

Query Match 77.5%; Score 31; DB 2; Length 372;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQIL 8
DB 296 LLLPLQVL 302

RESULT 43

hypothetical protein BH2276 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: D83934

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: D83934

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-392 <STO>

A:Cross-References: GB:AP001515; GB:BA000004; NID:g10174886; PIDN:BA05995.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2276

Query Match 77.5%; Score 31; DB 2; Length 392;
Best Local Similarity 87.5%; Pred. No. 1.7e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQIL 8
DB 290 LLLPLQTL 297

RESULT 44

C84306

hypothetical protein trp6 [imported] - Halobacterium sp. NRC-1

C:Species: Halobacterium sp. NRC-1

C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: C84306

R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S

; Leithausner, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jabl

Jung, K.H.; Alam, M.; Freitas, T.

Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000

A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Li

A:Title: Genome sequence of Halobacterium species NRC-1.

A:Reference number: A84160; MUID:20504483

A:Accession: C84306

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-447 <STO>

A:Cross-References: GB:AE004437; NID:g10581015; PIDN:AAG19815.1; GSPDB:GN00138

C:Genetics:

A:Gene: trp6

C:Superfamily: gamma-aminobutyric acid transporter

Query Match 77.5%; Score 31; DB 2; Length 447;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQIL 9
DB 376 VLLPLSVLL 384

RESULT 45

A71875

hypothetical protein jhp0880 - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C>Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: A71875

R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.

; Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F

Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric p

A:Reference number: A71800; MUID:99120557

A:Accession: A71875

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-493 <ARN>

A:Cross-References: GB:AE001518; GB:AE001439; NID:g4155454; PIDN:AAD06464.1; PID:g415

A:Experimental source: strain J99

C:Genetics:

A:Gene: jhp0880

Query Match 77.5%; Score 31; DB 2; Length 493;
Best Local Similarity 87.5%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
DB 254 LLLPLSILL 261

Search completed: November 6, 2002, 12:08:12
Job time : 14 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 8.11111 Seconds
(without alignments)
27.102 Million cell updates/sec

Title: US-09-905-083-32
Perfect score: 40
Sequence: 1 RLSSMVKKV 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	9	US-09-502-600-32	Sequence 32, Appl
2	40	100.0	154	US-09-261-416-7	Sequence 7, Appl
3	40	100.0	161	US-09-261-416-8	Sequence 8, Appl
4	40	100.0	224	US-08-944-483-33	Sequence 33, Appl
5	40	100.0	225	US-08-557-146-12	Sequence 12, Appl
6	40	100.0	225	US-09-027-337-4	Sequence 4, Appl
7	40	100.0	225	US-09-154-344-12	Sequence 12, Appl
8	40	100.0	253	US-08-557-146-2	Sequence 2, Appl
9	40	100.0	253	US-08-824-874-3	Sequence 3, Appl
10	40	100.0	253	US-09-154-344-2	Sequence 2, Appl
11	40	100.0	253	US-08-930-188-2	Sequence 2, Appl
12	40	100.0	253	US-09-210-084-3	Sequence 3, Appl
13	40	100.0	253	PCT-US96-04294-2	Sequence 2, Appl
14	36	90.0	9	US-09-502-600-110	Sequence 110, App
15	31	77.5	9	US-09-502-600-79	Sequence 79, Appl
16	31	77.5	9	US-09-502-600-87	Sequence 87, Appl
17	29	72.5	582	US-08-261-086-2	Sequence 2, Appl
18	29	72.5	582	US-08-261-086-4	Sequence 4, Appl
19	29	72.5	582	US-08-261-086-6	Sequence 6, Appl
20	28	70.0	459	US-09-491-785-2	Sequence 2, Appl
21	28	70.0	565	5171685-6	Patent No. 5171685
22	28	70.0	565	5118916-6	Patent No. 5118916
23	28	70.0	896	US-08-640-389A-10	Sequence 10, Appl
24	28	70.0	896	US-08-618-957A-10	Sequence 10, Appl
25	28	70.0	898	US-08-693-697-36	Sequence 36, Appl
26	28	70.0	906	US-08-640-389A-9	Sequence 9, Appl
27	28	70.0	906	US-08-618-957A-9	Sequence 9, Appl

28	70.0	908	2	US-08-693-697-33	Sequence 33, Appl
29	70.0	908	2	US-08-588-526-3	Sequence 3, Appl
30	70.0	958	2	US-08-640-389A-8	Sequence 8, Appl
31	70.0	958	4	US-08-618-957A-8	Sequence 8, Appl
32	70.0	960	1	US-08-355-888A-8	Sequence 8, Appl
33	70.0	960	2	US-08-588-190-3	Sequence 3, Appl
34	70.0	960	2	US-08-693-697-8	Sequence 8, Appl
35	70.0	960	2	US-08-640-389A-3	Sequence 3, Appl
36	70.0	960	3	US-08-693-696-8	Sequence 8, Appl
37	70.0	960	4	US-08-618-957A-3	Sequence 3, Appl
38	70.0	1165	2	US-08-640-389A-11	Sequence 11, Appl
39	70.0	1165	4	US-08-599-455B-4	Sequence 4, Appl
40	70.0	1165	4	US-09-093-814-1	Sequence 1, Appl
41	70.0	1165	4	US-09-069-781B-4	Sequence 4, Appl
42	70.0	1165	4	US-08-618-957A-11	Sequence 11, Appl
43	67.5	26	2	US-08-460-464B-18	Sequence 18, Appl
44	67.5	26	2	US-08-460-464B-20	Sequence 20, Appl
45	67.5	26	3	US-08-658-857B-9	Sequence 9, Appl

ALIGNMENTS

RESULT 1

US-09-502-600-32
; Sequence 32, Application US/09502600A
; Patent No. 6294344

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE OF INVENTION: Ovarian Cancer

; FILE REFERENCE: D6223CIP-C

; CURRENT FILING DATE: 2000-02-11

; CURRENT APPLICATION NUMBER: US/09/502,600A

; PRIOR APPLICATION NUMBER: 09/039,211

; PRIOR FILING DATE: 03-14-1998

; NUMBER OF SEQ ID NOS: 136

; SEQ ID NO 32

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Residues 123-131 of the SCE protein

US-09-502-600-32

Query Match 100.0%; Score 40; DB 4; Length 9;

Best Local Similarity 100.0%; Pred. No. 1.7e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 1 RLSSMVKKV 9

RESULT 2

US-09-261-416-7

; Sequence 7, Application US/09261416A

; Patent No. 6291663

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease

; FILE OF INVENTION: Overexpressed in Ovarian Carcinoma

; FILE REFERENCE: D6192

; CURRENT APPLICATION NUMBER: US/09/261,416A

; CURRENT FILING DATE: 1999-03-03

; NUMBER OF SEQ ID NOS: 14

; SEQ ID NO 7

; LENGTH: 154

; TYPE: PRT

; ORGANISM: Unknown

; FEATURE:

; OTHER INFORMATION: Serine protease catalytic domain of stratum corneum

; OTHER INFORMATION: chymotryptic enzyme (Scce) homologous to similar
; OTHER INFORMATION: domain in TAGD-12
US-09-261-416-7

Query Match 100.0%; Score 40; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.44;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
|||||
Db 69 RLSSMVKKV 77

RESULT 3

US-09-261-416-8
; Sequence 8, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 8
; LENGTH: 161
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of hepsin (Heps)
; OTHER INFORMATION: homologous to similar domain in TAGD-12
US-09-261-416-8

Query Match 100.0%; Score 40; DB 4; Length 161;
Best Local Similarity 100.0%; Pred. No. 0.46;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
|||||
Db 75 RLSSMVKKV 83

RESULT 4

US-08-944-483-33
; Sequence 33, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-33

Query Match 100.0%; Score 40; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
|||||
Db 94 RLSSMVKKV 102

RESULT 5

US-08-557-146-12
; Sequence 12, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide

US-08-557-146-12

Query Match 100.0%; Score 40; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.63; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
| | | | | | | | | |
DB 95 RLSSMVKKV 103

RESULT 6

US-09-027-337-4
; Sequence 4, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotsoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: TADG-15: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of Scce homologous to
; OTHER INFORMATION: similar domain in TADG-15
US-09-027-337-4

Query Match 100.0%; Score 40; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
| | | | | | | | | |
DB 95 RLSSMVKKV 103

RESULT 7

US-09-154-344-12
; Sequence 12, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESS: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
US-09-154-344-12

Query Match 100.0%; Score 40; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.63;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
| | | | | | | | | |
DB 95 RLSSMVKKV 103

RESULT 8

US-08-557-146-2
; Sequence 2, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESS: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-557-146-2

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
| | | | | | | | | |
DB 123 RLSSMVKKV 131

RESULT 9
US-08-824-874-3
; Sequence 3, Application US/08824874
; Patent No. 5962300
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE: Filed Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 532504
US-08-824-874-3

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
DB 123 RLSSMVKKV 131

RESULT 10
US-09-154-344-2
; Sequence 2, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-154-344-2

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
DB 123 RLSSMVKKV 131

RESULT 11
US-08-930-188-2
; Sequence 2, Application US/08930188
; Patent No. 6093397
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090

TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-930-188-2

Query Match 100.0%; Score 40; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMWKKV 9
DB 123 RLSSMWKKV 131

RESULT 12
US-09-210-084-3
Sequence 3, Application US/09210084
Patent No. 6197511
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL KALLIKREIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,084
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/824,874
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0252 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 532504
US-09-210-084-3

Query Match 100.0%; Score 40; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMWKKV 9
DB 123 RLSSMWKKV 131

RESULT 13
PCT-US96-04294-2
Sequence 2, Application PC/TUS9604294
GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Little, Sheila P.
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
RELATED NUCLEIC ACIDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/04294
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Blalock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-04294-2

Query Match 100.0%; Score 40; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.71;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMWKKV 9
DB 123 RLSSMWKKV 131

RESULT 14
US-09-502-600-110
Sequence 110, Application US/09502600A
Patent No. 6294344
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
Ovarian Cancer
FILE REFERENCE: D6223CIP-C
CURRENT FILING DATE: 2000-02-11
CURRENT APPLICATION NUMBER: US/09/502,600A
PRIOR FILING DATE: 09/039,211
PRIOR APPLICATION NUMBER: 03-14-1998
NUMBER OF SEQ ID NOS: 136
SEQ ID NO 110
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens

; FEATURE:
; OTHER INFORMATION: Residues 122-130 of the SCCE protein
US-09-502-600-110

Query Match 90.0%; Score 36; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
| | | | | | | |
Db 2 RLSSMVKK 9

RESULT 15

US-09-502-600-79
; Sequence 79, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 79
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 125-133 of the SCCE protein
US-09-502-600-79

Query Match 77.5%; Score 31; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SSMVKKV 9
| | | | | | | |
Db 1 SSMVKKV 7

RESULT 16

US-09-502-600-87
; Sequence 87, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 87
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 125-133 of the SCCE protein
US-09-502-600-87

Query Match 77.5%; Score 31; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 SSMVKKV 9
| | | | | | | |
Db 1 SSMVKKV 7

RESULT 17
US-08-261-086-2
; Sequence 2, Application US/08261086
; Patent No. 5539093
; GENERAL INFORMATION:
; APPLICANT: Fitzmaurice, Wayne P
; APPLICANT: Hellmann, Gary M
; APPLICANT: Grill, Laurence K
; APPLICANT: Kumagai, Monto H
; APPLICANT: della-Cioppa, Guy R
; TITLE OF INVENTION: DNA Sequences Encoding Enzymes Useful in
; TITLE OF INVENTION: Carotenoid Biosynthesis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: R. J. Reynolds Tobacco Company
; STREET: 950 Reynolds Blvd., P. O. Box 1487
; CITY: Winston-Salem
; STATE: No. 5539093th Carolina
; COUNTRY: U. S. A.
; ZIP: 27102-1487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,086
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Borschke, August J.
; REGISTRATION NUMBER: 30,539
; REFERENCE/DOCKET NUMBER: cc-212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 910-741-5491
; TELEFAX: 910-741-5449
; TELEX: NA
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 582 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-261-086-2

Query Match 72.5%; Score 29; DB 1; Length 582;
Best Local Similarity 55.6%; Pred. No. 2.4e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
| | | | | | | |
Db 343 RLNSRIKKI 351

RESULT 18

US-08-261-086-4
; Sequence 4, Application US/08261086
; Patent No. 5539093
; GENERAL INFORMATION:
; APPLICANT: Fitzmaurice, Wayne P
; APPLICANT: Hellmann, Gary M
; APPLICANT: Grill, Laurence K
; APPLICANT: Kumagai, Monto H
; APPLICANT: della-Cioppa, Guy R
; TITLE OF INVENTION: DNA Sequences Encoding Enzymes Useful in
; TITLE OF INVENTION: Carotenoid Biosynthesis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: R. J. Reynolds Tobacco Company
; STREET: 950 Reynolds Blvd., P. O. Box 1487
; CITY: Winston-Salem

; STATE: No. 5539093th Carolina
; COUNTRY: U. S. A.
; ZIP: 27102-1487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,086
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Borschke, August J.
; REGISTRATION NUMBER: 30,539
; REFERENCE/DOCKET NUMBER: cc-212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 910-741-5491
; TELEFAX: 910-741-5449
; TELEX: NA
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 582 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-261-086-4

Query Match 72.5%; Score 29; DB 1; Length 582;
Best Local Similarity 55.6%; Pred. No. 2.4e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||| :||:
Db 343 RLNSRIKKI 351

RESULT 19
US-08-261-086-6
; Sequence 6, Application US/08261086
; Patent No. 5539093
; GENERAL INFORMATION:
; APPLICANT: Fitzmaurice, Wayne P
; APPLICANT: Hellmann, Gary M
; APPLICANT: Grill, Laurence K
; APPLICANT: Kumagai, Monto H
; APPLICANT: della-Cioppa, Guy R
; TITLE OF INVENTION: DNA Sequences Encoding Enzymes Useful in
; TITLE OF INVENTION: Carotenoid Biosynthesis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: R. J. Reynolds Tobacco Company
; STREET: 950 Reynolds Blvd., P. O. Box 1487
; CITY: Winston-Salem
; STATE: No. 5539093th Carolina
; COUNTRY: U. S. A.
; ZIP: 27102-1487
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,086
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Borschke, August J.
; REGISTRATION NUMBER: 30,539
; REFERENCE/DOCKET NUMBER: cc-212
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 910-741-5491
; TELEFAX: 910-741-5449

; TELEX: NA
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 582 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-261-086-6

Query Match 72.5%; Score 29; DB 1; Length 582;
Best Local Similarity 55.6%; Pred. No. 2.4e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
||| :||:
Db 343 RLNSRIKKI 351

RESULT 20
US-09-491-785-2
; Sequence 2, Application US/09491785
; Patent No. 6316211
; GENERAL INFORMATION:
; APPLICANT: Zalacain, Magdalena
; APPLICANT: Biswas, Sanjoy
; APPLICANT: Burnham, Martin K. R.
; APPLICANT: Van Horn, Stephanie
; TITLE OF INVENTION: thdf
; FILE REFERENCE: GM10249
; CURRENT APPLICATION NUMBER: US/09/491,785
; CURRENT FILING DATE: 2000-01-26
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 459
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
; US-09-491-785-2

Query Match 70.0%; Score 28; DB 4; Length 459;
Best Local Similarity 62.3%; Pred. No. 3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
||| :||:
Db 155 RLSDLIKK 162

RESULT 21
5171685-6
; Patent No. 5171685
; APPLICANT: MCELWAIN, TERRY F.; HINES, STEPHEN A.; MCGUIRE,
; TRAVIS C.; PALMER, GUY H.; JASMER, DOUGLAS P.; REDUKER, DAVID W.
; GOFF, WILL L.; PERRYMAN, LANCE E.; DAVIS, WILLIAM C.
; TITLE OF INVENTION: CLONING OF THE BABESIA BOVIS 60 KD ANTIGEN
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/504,461
; FILING DATE: 04-APR-1990
; SEQ ID NO: 6;
; LENGTH: 565
5171685-6

Query Match 70.0%; Score 28; DB 6; Length 565;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
| : |||||
Db 277 RYADMVKKV 285

RESULT 22

5518916-6
; Patent No. 5518916
; APPLICANT: MCELWAIN, TERRY F.; MCGUIRE, TRAVIS C.; JASMER,
; DOUGLAS P.; REDUKER, DAVID W.; GOFF, WILL L.; STILLNER, DAVID
; TITLE OF INVENTION: CLONED BABESIA DNA
; NUMBER OF SEQUENCES: 7
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/342,480
; FILING DATE: 21-NOV-1994
; PRIORITY INFORMATION:
; APPLICATION NUMBER: 989,616
; FILING DATE: 14-DEC-1992
; APPLICATION NUMBER: 504,461
; FILING DATE: 04-APR-1990
; APPLICATION NUMBER: 333,155
; FILING DATE: 04-APR-1989
; SEQ ID NO: 6:
; LENGTH: 565
5518916-6

Query Match 70.0%; Score 28; DB 6; Length 565;
Best Local Similarity 66.7%; Pred. No. 3.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
I : |||||
DB 277 RYADVKKV 285

RESULT 23
US-08-640-389A-10
; Sequence 10, Application US/08640389A
; Patent No. 5912123
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: DETECTION OF THE LEPTIN
; TITLE OF INVENTION: RECEPTOR IN REPRODUCTIVE ORGANS AND METHODS FOR
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,389A
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 896 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide

US-08-640-389A-10

Query Match 70.0%; Score 28; DB 2; Length 896;
Best Local Similarity 75.0%; Pred. No. 5.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
I : || |||
DB 787 RISSSVKK 794

RESULT 24

US-08-618-957A-10
; Sequence 10, Application US/08618957A
; Patent No. 6355237
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas Joel
; APPLICANT: Shafer, Alan Wayne
; TITLE OF INVENTION: METHODS FOR USING THE OBESE
; TITLE OF INVENTION: GENE AND ITS GENE PRODUCT TO STIMULATE HEMATOPOIETIC
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,957A
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 008907-0033-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 896 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-618-957A-10

Query Match 70.0%; Score 28; DB 4; Length 896;
Best Local Similarity 75.0%; Pred. No. 5.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
I : || |||
DB 787 RISSSVKK 794

RESULT 25
US-08-693-697-36
; Sequence 36, Application US/08693697
; Patent No. 5869610

```
;
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cloffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/693,697
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0037-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 898 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-693-697-36
;
; Query Match 70.0%; Score 28; DB 2; Length 898;
; Best Local Similarity 75.0%; Pred. No. 5.7e+02;
; Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 RLSSVVK 8
; Db 789 RISSVVK 796
;
; RESULT 26
; US-08-640-389A-9
; Sequence 9, Application US/08640389A
; Patent No. 5912123
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cloffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: DETECTION OF THE LEPTIN
; TITLE OF INVENTION: RECEPTOR IN REPRODUCTIVE ORGANS AND METHODS FOR
; TITLE OF INVENTION: REGULATING REPRODUCTIVE BIOLOGY
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,389A
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 906 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
; US-08-640-389A-9
;
; Query Match 70.0%; Score 28; DB 2; Length 906;
; Best Local Similarity 75.0%; Pred. No. 5.8e+02;
; Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
;
; QY 1 RLSSVVK 8
; Db 787 RISSVVK 794
;
; RESULT 27
; US-08-618-957A-9
; Sequence 9, Application US/08618957A
; Patent No. 6355237
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Cloffi, Joseph
; APPLICANT: Zupancic, Thomas Joel
; APPLICANT: Shafer, Alan Wayne
; TITLE OF INVENTION: METHODS FOR USING THE OBESE
; TITLE OF INVENTION: GENE AND ITS GENE PRODUCT TO STIMULATE HEMATOPOIETIC
; TITLE OF INVENTION: DEVELOPMENT
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSEQ Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,957A
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 008907-0033-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
```

INFORMATION FOR SEQ ID NO: 9;
SEQUENCE CHARACTERISTICS:
LENGTH: 906 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-618-957A-9

Query Match 70.0%; Score 28; DB 4; Length 906;
Best Local Similarity 75.0%; Pred. No. 5.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSSVK 8
I:|I|I|I|
DB 789 RISSSVK 796

RESULT 28
US-08-693-697-33
Sequence 33, Application US/08693697
Patent No. 5869610
GENERAL INFORMATION:
APPLICANT: Snodgrass, H. R.
APPLICANT: Cioffi, Joseph
APPLICANT: Zupancic, Thomas J.
APPLICANT: Shafer, Alan W.
TITLE OF INVENTION: HU-B1.219, A NOVEL HUMAN HEMATOPOIETIN
NUMBER OF SEQUENCES: 38
TITLE OF INVENTION: RECEPTOR
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/693,697
FILING DATE: 05-AUG-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0037-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 33:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-693-697-33

Query Match 70.0%; Score 28; DB 2; Length 908;
Best Local Similarity 75.0%; Pred. No. 5.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSSVK 8
I:|I|I|I|
DB 789 RISSSVK 796

RESULT 29
US-08-588-526-3
Sequence 3, Application US/08588526
Patent No. 5882860
GENERAL INFORMATION:
APPLICANT: Snodgrass, H.
APPLICANT: Cioffi, Joseph
APPLICANT: Zupancic, Thomas
APPLICANT: Shafer, Alan
TITLE OF INVENTION: DETECTION OF A LEPTIN RECEPTOR
TITLE OF INVENTION: VARIANT
TITLE OF INVENTION: AND METHODS FOR REGULATING OBESITY
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: US
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/588,526
FILING DATE: 18-JAN-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-030
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 908 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-588-526-3

Query Match 70.0%; Score 28; DB 2; Length 908;
Best Local Similarity 75.0%; Pred. No. 5.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSSVK 8
I:|I|I|I|
DB 789 RISSSVK 796

RESULT 30
US-08-640-389A-8
Sequence 8, Application US/08640389A
Patent No. 5912123
GENERAL INFORMATION:
APPLICANT: Snodgrass, H. R.
APPLICANT: Cioffi, Joseph
APPLICANT: Zupancic, Thomas J.
APPLICANT: Shafer, Alan W.
TITLE OF INVENTION: DETECTION OF THE LEPTIN
TITLE OF INVENTION: RECEPTOR IN REPRODUCTIVE ORGANS AND METHODS FOR
TITLE OF INVENTION: REGULATING REPRODUCTIVE BIOLOGY
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA

Query Match 70.0%; Score 28; DB 2; Length 908;
Best Local Similarity 75.0%; Pred. No. 5.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSSVK 8
I:|I|I|I|
DB 789 RISSSVK 796

ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/640,389A
FILING DATE: 29-APR-1996
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-032
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 958 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-640-389A-8

Query Match 70.0%; Score 28; DB 2; Length 958;
Best Local Similarity 75.0%; Pred. No. 6.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
I:| | | | |
DB 787 RISSSVKK 794

RESULT 31
US-08-618-957A-8
; Sequence 8, Application US/08618957A
; Patent No. 6355237
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas Joel
; APPLICANT: Shafer, Alan Wayne
; TITLE OF INVENTION: METHODS FOR USING THE OBSE
; TITLE OF INVENTION: GENE AND ITS GENE PRODUCT TO STIMULATE HEMATOPOIETIC
; DEVELOPMENT
; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/618,957A
FILING DATE: 20-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 008907-0033-999
TELECOMMUNICATION INFORMATION:

TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 958 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-618-957A-8

Query Match 70.0%; Score 28; DB 4; Length 958;
Best Local Similarity 75.0%; Pred. No. 6.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
I:| | | | |
DB 787 RISSSVKK 794

RESULT 32
US-08-355-888A-8
; Sequence 8, Application US/08355888A
; Patent No. 5763211
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: HU-B1.219, A NOVEL HUMAN HEMATOPOIETIN
; RECEPTOR
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC Compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/355,888A
FILING DATE: 14-DEC-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 7225-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 960 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-355-888A-8

Query Match 70.0%; Score 28; DB 1; Length 960;
Best Local Similarity 75.0%; Pred. No. 6.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
I:| | | | |
DB 789 RISSSVKK 796

RESULT 33
US-08-588-190-3
; Sequence 3, Application US/08588190
; Patent No. 5856098
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas Joel
; APPLICANT: Shafer, Alan Wayne
; TITLE OF INVENTION: DETECTION OF A LEPTIN RECEPTOR
; TITLE OF INVENTION: VARIANT AND METHODS FOR REGULATING OBESITY
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/588,190
; FILING DATE: 18-JAN-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 008907-0029-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 960 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-588-190-3

Query Match 70.0%; Score 28; DB 2; Length 960;
Best Local Similarity 75.0%; Pred. No. 6.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
Db 789 RISSSVKK 796

RESULT 34
US-08-693-697-8
; Sequence 8, Application US/08693697
; Patent No. 5869610
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
; TITLE OF INVENTION: RECEPTOR
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of The Americas

; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: FastSeq for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/693,697
; FILING DATE: 05-AUG-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0037-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 960 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-693-697-8

Query Match 70.0%; Score 28; DB 2; Length 960;
Best Local Similarity 75.0%; Pred. No. 6.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSMVKK 8
Db 789 RISSSVKK 796

RESULT 35
US-08-640-389A-3
; Sequence 3, Application US/08640389A
; Patent No. 5912123
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. R.
; APPLICANT: Cioffi, Joseph
; APPLICANT: Zupancic, Thomas J.
; APPLICANT: Shafer, Alan W.
; TITLE OF INVENTION: DETECTION OF THE LEPTIN
; TITLE OF INVENTION: RECEPTOR IN REPRODUCTIVE ORGANS AND METHODS FOR
; TITLE OF INVENTION: REGULATING REPRODUCTIVE BIOLOGY
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,389A
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 960 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-640-389A-3

Query Match 70.0%; Score 28; DB 2; Length 960;
Best Local Similarity 75.0%; Pred. No. 6.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSSVKK 8
|:|:| |

Db 789 RISSSVKK 796

RESULT 36
US-08-693-696-8
Sequence 8, Application US/08693696
Patent No. 6005080
GENERAL INFORMATION:
APPLICANT: Snodgrass, H. R.
APPLICANT: Cioffi, Joseph
APPLICANT: Zupancic, Thomas J.
APPLICANT: Shafer, Alan W.
TITLE OF INVENTION: Hu-B1.219, A NOVEL HUMAN HEMATOPOIETIN
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/693,696
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/355,888
FILING DATE: 14-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 7225-078
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 960 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-693-696-8

Query Match 70.0%; Score 28; DB 3; Length 960;
Best Local Similarity 75.0%; Pred. No. 6.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSSVKK 8
|:|:| |

Db 789 RISSSVKK 796

RESULT 37
US-08-618-957A-3
Sequence 3, Application US/08618957A
Patent No. 6355237
GENERAL INFORMATION:
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Cioffi, Joseph
APPLICANT: Zupancic, Thomas Joel
APPLICANT: Shafer, Alan Wayne
TITLE OF INVENTION: METHODS FOR USING THE OBESE
TITLE OF INVENTION: GENE AND ITS GENE PRODUCT TO STIMULATE HEMATOPOIETIC
TITLE OF INVENTION: DEVELOPMENT
NUMBER OF SEQUENCES: 28
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of The Americas
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036-2811
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/618,957A
FILING DATE: 20-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 008907-0033-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-493-4935
TELEFAX: 650-493-5556
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 960 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-618-957A-3

Query Match 70.0%; Score 28; DB 4; Length 960;
Best Local Similarity 75.0%; Pred. No. 6.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RLSSSVKK 8
|:|:| |

Db 789 RISSSVKK 796

RESULT 38
US-08-640-389A-11
Sequence 11, Application US/08640389A
Patent No. 5912123
GENERAL INFORMATION:
APPLICANT: Snodgrass, H. R.
APPLICANT: Cioffi, Joseph
APPLICANT: Zupancic, Thomas J.
APPLICANT: Shafer, Alan W.
TITLE OF INVENTION: DETECTION OF THE LEPTIN
TITLE OF INVENTION: RECEPTOR IN REPRODUCTIVE BIOLOGY
TITLE OF INVENTION: REGULATING REPRODUCTIVE BIOLOGY
NUMBER OF SEQUENCES: 16

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/640,389A
; FILING DATE: 29-APR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-032
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1165 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-640-389A-11

Query Match 70.0%; Score 28; DB 2; Length 1165;
Best Local Similarity 75.0%; Pred. No. 7.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLSSMVKK 8
I:|I| I|I|
Db 787 RISSSVKK 794

RESULT 39
US-08-599-455B-4
; Sequence 4, Application US/08599455B
; Patent No. 5972621
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Tepper, Robert I.
; TITLE OF INVENTION: METHODS OF IDENTIFYING COMPOUNDS THAT
; TITLE OF INVENTION: MODULATE BODY WEIGHT USING THE OB RECEPTOR
; NUMBER OF SEQUENCES: 44
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/599,455B
; FILING DATE: 22-JAN-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/583,153
; FILING DATE: 28-DEC-1995
; APPLICATION NUMBER: 08/570,142
; FILING DATE: 11-DEC-1995

; APPLICATION NUMBER: 08/569,485
; FILING DATE: 08-DEC-1995
; APPLICATION NUMBER: 08/566,622
; FILING DATE: 04-DEC-1995
; APPLICATION NUMBER: 08/562,663
; FILING DATE: 27-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Meiklejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/017001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1165 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-599-455B-4

Query Match 70.0%; Score 28; DB 2; Length 1165;
Best Local Similarity 75.0%; Pred. No. 7.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLSSMVKK 8
I:|I| I|I|
Db 787 RISSSVKK 794

RESULT 40
US-09-093-814-1
; Sequence 1, Application US/09093814
; Patent No. 6270981
; GENERAL INFORMATION:
; APPLICANT: Carpenter et al.
; TITLE OF INVENTION: ASSAY SYSTEMS FOR LEPTIN-ENHANCING AGENTS
; FILE REFERENCE: REG 580-A
; CURRENT APPLICATION NUMBER: US/09/093,814
; CURRENT FILING DATE: 1998-06-09
; PRIOR APPLICATION NUMBER: 60/049,108
; PRIOR FILING DATE: 1997-06-09
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1165
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-093-814-1

Query Match 70.0%; Score 28; DB 4; Length 1165;
Best Local Similarity 75.0%; Pred. No. 7.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RLSSMVKK 8
I:|I| I|I|
Db 787 RISSSVKK 794

RESULT 41
US-09-069-781B-4
; Sequence 4, Application US/09069781B
; Patent No. 6287782
; GENERAL INFORMATION:
; APPLICANT: Tartaglia, Louis A.
; APPLICANT: Tepper, Robert I.
; APPLICANT: Culpepper, Janice A.
; APPLICANT: White, David W.
; TITLE OF INVENTION: THE OB RECEPTOR AND METHODS FOR
; TITLE OF INVENTION: THE DIAGNOSIS AND TREATMENT OF BODY WEIGHT DISORDERS,
; TITLE OF INVENTION: INCLUDING OBESITY AND CACHEXIA

```

; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US 09/069,781B
; FILING DATE: 29-APRIL-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/864,564
; FILING DATE: 28-MAY-1997
; APPLICATION NUMBER: US 08/708,123
; FILING DATE: 03-SEP-1996
; APPLICATION NUMBER: US 08/638,524
; FILING DATE: 26-APR-1996
; APPLICATION NUMBER: US 08/599,455
; FILING DATE: 22-JAN-1996
; APPLICATION NUMBER: US 08/583,153
; FILING DATE: 28-DEC-1995
; APPLICATION NUMBER: US 08/570,142
; FILING DATE: 11-DEC-1995
; APPLICATION NUMBER: US 08/569,485
; FILING DATE: 08-DEC-1995
; APPLICATION NUMBER: US 08/566,622
; FILING DATE: 04-DEC-1995
; APPLICATION NUMBER: US 08/562,663
; FILING DATE: 27-NOV-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkielejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 07334/082001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 542-5070
; TELEFAX: (617) 542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1165 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; FRAGMENT TYPE: Internal
; US-09-069-781B-4

```

Query Match 70.0%; Score 28; DB 4; Length 1165;
 Best Local Similarity 75.0%; Pred. No. 7.4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 RLSSMVKK 8
   |::| |::|
DB 787 RISSSVKK 794

```

```

RESULT 42
US-08-618-957A-11
; Sequence 11, Application US/08618957A
; Patent No. 6355237
; GENERAL INFORMATION:
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Clotfii, Joseph
; APPLICANT: Zupancic, Thomas Joel
; APPLICANT: Shafer, Alan Wayne
; TITLE OF INVENTION: METHODS FOR USING THE OBESE
; TITLE OF INVENTION: GENE AND ITS GENE PRODUCT TO STIMULATE HEMATOPOIETIC
; TITLE OF INVENTION: DEVELOPMENT

```

```

; NUMBER OF SEQUENCES: 28
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of The Americas
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036-2811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/618,957A
; FILING DATE: 20-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 008907-0033-999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-493-4935
; TELEFAX: 650-493-5556
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1165 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-618-957A-11

```

Query Match 70.0%; Score 28; DB 4; Length 1165;
 Best Local Similarity 75.0%; Pred. No. 7.4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

QY 1 RLSSMVKK 8
   |::| |::|
DB 787 RISSSVKK 794

```

```

RESULT 43
US-08-460-464B-18
; Sequence 18, Application US/08460464B.
; Patent No. 5877274
; GENERAL INFORMATION:
; APPLICANT: Hancock, Robert E.W.
; APPLICANT: Karunaratne, Nedra
; TITLE OF INVENTION: ANTIMICROBIAL CATIONIC PEPTIDES
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/460,464B
; FILING DATE: 02-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:

```

FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Haile, Ph.D., Lisa A.
 REGISTRATION NUMBER: 38,347
 REFERENCE/DOCKET NUMBER: 07420/008001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619/678-5070
 TELEFAX: 619/678-5099
 INFORMATION FOR SEQ ID NO: 18:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 26 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-460-464B-18

Query Match 67.5%; Score 27; DB 2; Length 26;
 Best Local Similarity 55.6%; Pred. No. 30;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 :|:|:|:|
 Db 8 KLTSVLKKV 16

RESULT 44
 US-08-460-464B-20
 Sequence 20, Application US/08460464B
 Patent No. 5877274
 GENERAL INFORMATION:
 APPLICANT: Hancock, Robert E.W.
 APPLICANT: Karunaratne, Nedra
 TITLE OF INVENTION: ANTIMICROBIAL CATIONIC PEPTIDES
 NUMBER OF SEQUENCES: 21
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson, P.C.
 STREET: 4225 Executive Square, Suite 1400
 CITY: La Jolla
 STATE: CA
 COUNTRY: US
 ZIP: 92037

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/460,464B
 FILING DATE: 02-JUN-1995
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Haile, Ph.D., Lisa A.
 REGISTRATION NUMBER: 38,347
 REFERENCE/DOCKET NUMBER: 07420/008001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619/678-5070
 TELEFAX: 619/678-5099
 INFORMATION FOR SEQ ID NO: 20:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 26 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-460-464B-20

Query Match 67.5%; Score 27; DB 2; Length 26;
 Best Local Similarity 55.6%; Pred. No. 30;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9

Db 8 KLTSVLKKV 16
 :|:|:|:|

RESULT 45
 US-08-658-857B-9
 Sequence 9, Application US/08658857B
 Patent No. 6040435
 GENERAL INFORMATION:
 APPLICANT: Hancock, Robert E. W.
 APPLICANT: Karunaratne, Nedra
 TITLE OF INVENTION: ANTIMICROBIAL CATIONIC PEPTIDES
 NUMBER OF SEQUENCES: 37
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Fish & Richardson P.C.
 STREET: 4225 Executive Square, Suite 1400
 CITY: La Jolla
 STATE: CA
 COUNTRY: USA
 ZIP: 92037

COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: Windows95
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/658,857B
 FILING DATE: May 31, 1996
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/460,464
 FILING DATE: June 2, 1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Haile, Lisa A.
 REGISTRATION NUMBER: 38,347
 REFERENCE/DOCKET NUMBER: 07420/014001
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 619/678-5070
 TELEFAX: 619/678-5099
 INFORMATION FOR SEQ ID NO: 9:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 26 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-658-857B-9

Query Match 67.5%; Score 27; DB 3; Length 26;
 Best Local Similarity 55.6%; Pred. No. 30;
 Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 RLSSMVKKV 9
 :|:|:|:|
 Db 8 KLTSVLKKV 16

Search completed: November 6, 2002, 12:06:21
 Job time : 9.11111 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 8.11111 Seconds
(without alignments)
27.102 Million cell updates/sec

Title: US-09-905-083-33
Perfect score: 40
Sequence: 1 LLLPLQILL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2.6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2.6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2.6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2.6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2.6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2.6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	9	4	US-09-502-600-33
2	40	100.0	253	2	US-08-557-146-2
3	40	100.0	253	2	US-08-824-874-3
4	40	100.0	253	2	US-09-154-344-2
5	40	100.0	253	3	US-08-930-188-2
6	40	100.0	253	4	US-09-210-084-3
7	40	100.0	253	5	PCT-US96-04294-2
8	36	90.0	9	4	US-09-502-600-35
9	36	90.0	9	4	US-09-502-600-36
10	33	82.5	190	1	US-08-339-152A-19
11	33	82.5	190	2	US-08-007-999B-6
12	33	82.5	190	2	US-08-689-276A-6
13	33	82.5	634	1	US-08-339-152A-17
14	33	82.5	653	1	US-08-339-152A-16
15	33	82.5	653	2	US-08-007-999B-3
16	33	82.5	653	2	US-08-689-276A-3
17	32	80.0	21	3	US-08-753-007A-24
18	32	80.0	21	4	US-09-398-496-24
19	32	80.0	22	4	US-08-977-378-22
20	32	80.0	235	1	US-07-940-605A-12
21	32	80.0	235	2	US-08-690-096-12
22	32	80.0	376	3	US-08-751-512-8
23	32	80.0	492	1	US-07-794-393-4
24	32	80.0	492	1	US-08-001-711-4
25	32	80.0	1306	4	US-08-989-299-7
26	31	77.5	270	2	US-08-773-368-1
27	31	77.5	270	3	US-09-199-887-1

28	77.5	445	4	US-08-974-691-6	Sequence 6, Appli
29	77.5	451	4	US-08-974-691-2	Sequence 2, Appli
30	77.5	625	3	US-08-996-139-15	Sequence 15, Appli
31	77.5	625	4	US-08-995-659-15	Sequence 15, Appli
32	77.5	625	4	US-09-215-649A-15	Sequence 15, Appli
33	77.5	629	1	US-08-278-635B-6	Sequence 6, Appli
34	77.5	629	3	US-08-464-258B-6	Sequence 6, Appli
35	77.5	629	3	US-08-471-961-6	Sequence 6, Appli
36	77.5	673	4	US-09-063-950-2	Sequence 2, Appli
37	77.5	699	1	US-08-348-006B-7	Sequence 7, Appli
38	77.5	699	2	US-08-800-825A-7	Sequence 7, Appli
39	77.5	699	4	US-09-158-657-7	Sequence 7, Appli
40	77.5	702	4	US-09-232-200-102	Sequence 102, App
41	77.5	702	4	US-09-232-197-102	Sequence 102, App
42	77.5	702	4	US-09-232-201-102	Sequence 102, App
43	77.5	990	4	US-09-627-376-7	Sequence 7, Appli
44	77.5	4302	3	US-08-658-136-5	Sequence 5, Appli
45	77.5	4303	2	US-08-460-751-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-502-600-33
; Sequence 33, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US/09/502,600A
; PRIOR FILING DATE: 09/039,211
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 33
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 5-13 of the SCCE protein
US-09-502-600-33

Query Match 100.0%; Score 40; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	LLLPLQILL	9
Db	1	LLLPLQILL	9

RESULT 2
US-08-557-146-2
; Sequence 2, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,146
FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-181
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-557-146-2

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
| | | | | | | | |
Db 5 LLLPLQILL 13

RESULT 3
US-08-824-874-3
Sequence 3, Application US/08824874
Patent No. 5962300
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL KALLIKREIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/824,874
FILING DATE: Filed Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0252 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:

LIBRARY: GenBank
CLONE: 532504
US-08-824-874-3
Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLLPLQILL 9
| | | | | | | | |
Db 5 LLLPLQILL 13

RESULT 4
US-09-154-344-2
Sequence 2, Application US/09154344
Patent No. 5981256
GENERAL INFORMATION:
APPLICANT: Egelrud, Torbjorn
APPLICANT: Hansson, Lennart
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
TITLE OF INVENTION: Enzyme (SCCE)
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case, Patent Department
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,344
FILING DATE: 16-SEP-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/557,146
FILING DATE: 14-DEC-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-181
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-154-344-2

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
| | | | | | | | |
Db 5 LLLPLQILL 13

RESULT 5
US-08-930-188-2
Sequence 2, Application US/08930188
Patent No. 6093397
GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.


```

; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,188
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-930-188-2

Query Match 100.0%; Score 40; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 5 LLLPLQILL 13

RESULT 6
US-09-210-084-3
; Sequence 3, Application US/09210084
; Patent No. 6197511
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,084
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 532504
; US-09-210-084-3

Query Match 100.0%; Score 40; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 5 LLLPLQILL 13

RESULT 7
PCT-US96-04294-2
; Sequence 2, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
PCT-US96-04294-2

Query Match 100.0%; Score 40; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||
Db 5 LLLPLQILL 13

RESULT 8

US-09-502-600-35
; Sequence 35, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 35
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 6-14 of the SCCE protein
US-09-502-600-35

Query Match 90.0%; Score 36; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
|||||
Db 1 LLLPLQILL 8

RESULT 9

US-09-502-600-36
; Sequence 36, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 36
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 4-12 of the SCCE protein
US-09-502-600-36

Query Match 90.0%; Score 36; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 8
|||||
Db 2 LLLPLQILL 9

Query Match 100.0%; Score 40; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||
Db 5 LLLPLQILL 13

US-09-502-600-35
; Sequence 35, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 35
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 6-14 of the SCCE protein
US-09-502-600-35

Query Match 90.0%; Score 36; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
|||||
Db 1 LLLPLQILL 8

US-09-502-600-36
; Sequence 36, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 36
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 4-12 of the SCCE protein
US-09-502-600-36

Query Match 90.0%; Score 36; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 8
|||||
Db 2 LLLPLQILL 9

RESULT 10

US-08-339-152A-19
; Sequence 19, Application US/08339152A
; Patent No. 5643726
; GENERAL INFORMATION:
; APPLICANT: Tanzi, Rudolph E.
; APPLICANT: Kovacs, Dora M.
; TITLE OF INVENTION: Methods For Modulating Transcription
; TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339.152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 0609.4120000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; TELEX:
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-339-152A-19

Query Match 82.5%; Score 33; DB 1; Length 190;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||
Db 1 LLLPLQILL 9

RESULT 11

US-08-007-999B-6
; Sequence 6, Application US/08007999B
; Patent No. 5851787
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/007,999B
FILING DATE: 21-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,642
FILING DATE: 20-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,022
FILING DATE: 17-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, G. Kevin
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT
TELEPHONE: (202)371-2571
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 190 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-007-999B-6

Query Match 82.5%; Score 33; DB 2; Length 190;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 ||||| :||
DB 1 LLLPLSLLL 9

RESULT 12
US-08-276A-6
; Sequence 6, Application US/08689276A
; Patent No. 5891991
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzil, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,276A
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:

NAME: JORGE A. GOLDSTEIN
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 0609.3520003
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2571
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 190 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-689-276A-6

Query Match 82.5%; Score 33; DB 2; Length 190;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 ||||| :||
DB 1 LLLPLSLLL 9

RESULT 13
US-08-339-152A-17
; Sequence 17, Application US/08339152A
; Patent No. 5643726
; GENERAL INFORMATION:
; APPLICANT: Tanzil, Rudolph E.
; APPLICANT: Kovacs, Dora M.
; TITLE OF INVENTION: Methods For Modulating Transcription
; TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 0609.4120000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; TELEX:

INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 634 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-339-152A-17

Query Match 82.5%; Score 33; DB 1; Length 634;
Best Local Similarity 77.8%; Pred. No. 82;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 ||||| :||
DB 1 LLLPLSLLL 9

```
RESULT 14
US-08-339-152A-16
; Sequence 16, Application US/08339152A
; Patent No. 5643726
; GENERAL INFORMATION:
; APPLICANT: Tanzi, Rudolph E.
; APPLICANT: Kovacs, Dora M.
; TITLE OF INVENTION: Methods For Modulating Transcription
; TITLE OF INVENTION: From The Amyloid -protein Precursor (APP) Promoter
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 0609.4120000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; TELEX:
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 653 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-339-152A-16

Query Match 82.5%; Score 33; DB 1; Length 653;
Best Local Similarity 77.8%; Pred. No. 85;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 21 LLLPLSLLL 29

RESULT 15
US-08-007-999B-3
; Sequence 3, Application US/08007999B
; Patent No. 5851787
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,276A
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
```

```
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/007,999B
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, G. Kevin
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 653 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-007-999B-3

Query Match 82.5%; Score 33; DB 2; Length 653;
Best Local Similarity 77.8%; Pred. No. 85;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 21 LLLPLSLLL 29

RESULT 16
US-08-689-276A-3
; Sequence 3, Application US/08689276A
; Patent No. 5891991
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,276A
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
```

ATTORNEY/AGENT INFORMATION:
NAME: JORGE A. GOLDSTEIN
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 0609.3520003
TELEPHONE: (202)371-2571
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 653 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-689-276A-3

Query Match 82.5%; Score 33; DB 2; Length 653;
Best Local Similarity 77.8%; Pred. No. 85;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
||||| :||
DB 21 LLLPLSLLL 29

RESULT 17
US-08-753-007A-24
Sequence 24, Application US/08753007A
Patent No. 6074841
GENERAL INFORMATION:
APPLICANT: Gearing, David P.
APPLICANT: Busfield, Samantha J.
TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/753,007A
FILING DATE: 19-NOV-1996
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/699,591
FILING DATE: 19-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07334/022001
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX:
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-753-007A-24

Query Match 80.0%; Score 32; DB 3; Length 21;
Best Local Similarity 77.8%; Pred. No. 3.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
||||| :||
DB 8 LLLPLALLL 16

RESULT 18
US-09-398-496-24
Sequence 24, Application US/09398496
Patent No. 6133423
GENERAL INFORMATION:
APPLICANT: Gearing, David P.
APPLICANT: Busfield, Samantha J.
TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
TITLE OF INVENTION: AND USES THEREFOR
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSEQ Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/398,496
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/753,007
FILING DATE: 19-NOV-1996
APPLICATION NUMBER: 08/699,591
FILING DATE: 19-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Fasse, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 07334/022001
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX:
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 21 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-398-496-24

Query Match 80.0%; Score 32; DB 4; Length 21;
Best Local Similarity 77.8%; Pred. No. 3.3;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
||||| :||
DB 8 LLLPLALLL 16

RESULT 19
US-08-977-378-22
Sequence 22, Application US/08977378
Patent No. 6309842
GENERAL INFORMATION:
APPLICANT: Dower, William J.
Gates, Christopher M.
Heinkel, Gregory L.
Lalonde, Guy
Mattheakis, Larry C.

;; Paddon, Christopher J.
;; Schatz, Peter J.
;; TITLE OF INVENTION: Use of Modified Tethers in Screening
;; Compound Libraries
;;
;; NUMBER OF SEQUENCES: 31
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Townsend and Townsend and Crew LLP
;; STREET: Two Embarcadero Center, Eighth Floor
;; CITY: San Francisco
;; STATE: California
;; COUNTRY: USA
;; ZIP: 94111-3834
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/977,378
;; FILING DATE: 24-NOV-1997
;; CLASSIFICATION: <Unknown>
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/758,307
;; FILING DATE: 03-DEC-1996
;;
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Liebeschuetz, Joe
;; REGISTRATION NUMBER: 37,505
;; REFERENCE/DOCKET NUMBER: 16528A-018010US
;;
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;;
;; INFORMATION FOR SEQ ID NO: 22:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 22 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: <Unknown>
;; TOPOLOGY: linear
;;
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-08-977-378-22

Query Match 80.0%; Score 32; DB 4; Length 22;
Best Local Similarity 77.8%; Pred. NO. 3.5;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLLPLQILL 9
Db 8 LLLPLALLL 16

RESULT 20
US-07-940-605A-12
; Sequence 12, Application US/07940605A
; Patent No. 5340926
; GENERAL INFORMATION:
; APPLICANT: ARUFFO, ALEJANDRO
; APPLICANT: HOLLENBAUGH, DIANE
; APPLICANT: LEDBETTER, JEFFREY A.
; TITLE OF INVENTION: SOLUBLE LIGANDS FOR CD40
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

;; APPLICATION NUMBER: US/07/940,605A
;; FILING DATE: 04-SEP-1992
;; CLASSIFICATION: 530
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Misrock, S. Leslie
;; REGISTRATION NUMBER: 18,872
;; REFERENCE/DOCKET NUMBER: 5624-184
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 212 790-9090
;; TELEFAX: 212 869-8864/9741
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 235 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-07-940-605A-12

Query Match 80.0%; Score 32; DB 1; Length 235;
Best Local Similarity 77.8%; Pred. NO. 44;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLLPLQILL 9
Db 8 LLLPLALLL 16

RESULT 21
US-08-690-096-12
; Sequence 12, Application US/08690096
; Patent No. 5945513
; GENERAL INFORMATION:
; APPLICANT: ARUFFO, ALEJANDRO
; APPLICANT: HOLLENBAUGH, DIANE
; APPLICANT: LEDBETTER, JEFFREY A.
; TITLE OF INVENTION: SOLUBLE LIGANDS FOR CD40
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/690,096
; FILING DATE: 31-JUL-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/940,605
; FILING DATE: 04-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 5624-184
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-690-096-12

Query Match 80.0%; Score 32; DB 2; Length 235;
Best Local Similarity 77.8%; Pred. No. 44;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
| | | | | : | |
Db 8 LLLPLALLL 16

RESULT 22

US-08-751-512-8
; Sequence 8, Application US/08/751512
; Patent No. 6001962
; GENERAL INFORMATION:
; APPLICANT: Ramey, J. Kevin
; APPLICANT: Williams, Lewis T.
; TITLE OF INVENTION: Modified PAS Ligands
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy, Matthew B.
; REGISTRATION NUMBER: 39,787
; REFERENCE/DOCKET NUMBER: 02307K-07100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-512-8

Query Match 80.0%; Score 32; DB 3; Length 376;
Best Local Similarity 77.8%; Pred. No. 73;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
| | | | | : | |
Db 8 LLLPLALLL 16

RESULT 23

US-07-794-393-4
; Sequence 4, Application US/07794393
; Patent No. 5236844
; GENERAL INFORMATION:
; APPLICANT: CHAMBER, PIERRE
; APPLICANT: BASSET, PAUL
; APPLICANT: BELLOCQ, JEAN-PIERRE
; TITLE OF INVENTION: ANALYTICAL MARKERS FOR MALIGNANT BREAST
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Ave. NW Suite 300
; CITY: Washington

STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/794,393
FILING DATE: 19911121
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9025326.1
FILING DATE: 21-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: GOLDSTEIN, JORGE A
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 1383.0040000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 466-0800
TELEFAX: (202) 833-8716
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 492 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-794-393-4

Query Match 80.0%; Score 32; DB 1; Length 492;
Best Local Similarity 77.8%; Pred. No. 97;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
| | | | | : | |
Db 15 LLLPLPLLL 23

RESULT 24

US-08-001-711-4
; Sequence 4, Application US/08001711
; Patent No. 5484726
; GENERAL INFORMATION:
; APPLICANT: BASSET, PAUL
; APPLICANT: BELLOCQ, JEAN-PIERRE
; APPLICANT: CHAMBER, PIERRE
; TITLE OF INVENTION: ANALYTICAL MARKERS FOR MALIGNANT BREAST
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Suite 300
; CITY: Washington
; STATE: D.C.
; ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/001,711
FILING DATE: 19930107
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/794,393
FILING DATE: 11-NOV-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9025626.1
FILING DATE: 21-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: MILLMAN, ROBERT A

```

; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 1383.0040001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)466-0800
; TELEFAX: (202)833-8716
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-001-711-4

Query Match      80.0%; Score 32; DB 1; Length 492;
Best Local Similarity 77.8%; Pred. No. 97;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 15 LLLPLPLL 23

RESULT 25
US-08-989-299-7
; Sequence 7, Application US/08989299
; Patent No. 6194556
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
; TITLE OF INVENTION: AND THERAPEUTIC AND DIAGNOSTIC USES THEREFOR
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,299
; FILING DATE: 11-DEC-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold E., Beth
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: MIA-025.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1306 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-989-299-7

Query Match      80.0%; Score 32; DB 4; Length 1306;
Best Local Similarity 77.8%; Pred. No. 2.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 12 LLLPLPLL 20
```

```

RESULT 26
US-08-773-368-1
; Sequence 1, Application US/08773368
; Patent No. 5856130
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/773,368
; FILING DATE:
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy RJ
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0186 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 270 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: 1599164
US-08-773-368-1

Query Match      77.5%; Score 31; DB 2; Length 270;
Best Local Similarity 77.8%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 10 LLLPLLLL 18

RESULT 27
US-09-199-887-1
; Sequence 1, Application US/09199887
; Patent No. 6071874
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Goli, Surya K.
; TITLE OF INVENTION: NOVEL HUMAN PATHOGENESIS-RELATED
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
```


COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/199,887
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/773,368
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy RJ
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0186 US
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 270 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: 1599164
US-09-199-887-1

Query Match 77.5%; Score 31; DB 3; Length 270;
Best Local Similarity 77.8%; Pred. No. 79;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPQILL 9
Db 10 LLLPLLLL 18

RESULT 28
US-08-974-691-6
Sequence 6, Application US/08974691
Patent No. 6225103
GENERAL INFORMATION:
APPLICANT: Keolsch, Gerald
APPLICANT: Lin, Xinli
APPLICANT: Tang, Jordan
TITLE OF INVENTION: Cloning and Characterization of Napsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 W. Peachtree
STREET: St.
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,691
FILING DATE: 20-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,196
FILING DATE: 20-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,126
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRF 166
TELEPHONE: 404-873-8794
TELEFAX: 404-873-8795
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 451 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRF 166
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-873-8794
TELEFAX: 404-873-8795
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 445 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-974-691-6

Query Match 77.5%; Score 31; DB 4; Length 445;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPQILL 9
Db 6 LLLPLLLL 14

RESULT 29
US-08-974-691-2
Sequence 2, Application US/08974691
Patent No. 6225103
GENERAL INFORMATION:
APPLICANT: Keolsch, Gerald
APPLICANT: Lin, Xinli
APPLICANT: Tang, Jordan
TITLE OF INVENTION: Cloning and Characterization of Napsin
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 W. Peachtree
STREET: St.
CITY: Atlanta
STATE: GA
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,691
FILING DATE: 20-NOV-1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/031,196
FILING DATE: 20-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046,126
FILING DATE: 09-MAY-1997
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRF 166
TELEPHONE: 404-873-8794
TELEFAX: 404-873-8795
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 451 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein

ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-08-974-691-2

Query Match 77.5%; Score 31; DB 4; Length 451;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||:||
Db 6 LLLPLLLLL 14

RESULT 30

US-08-996-139-15
Sequence 15, Application US/08996139
Patent No. 6017729

GENERAL INFORMATION:
APPLICANT: Anderson, Dirk M.
APPLICANT: Galibert, Laurent
APPLICANT: Maraskovsky, Eugene
TITLE OF INVENTION: Receptor Activator of NF-kappaB
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation, Law Department
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Power Macintosh 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/996,139
FILING DATE: 22 DECEMBER 1997

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/064,671
FILING DATE: 14 OCTOBER 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/813,509
FILING DATE: 07 MARCH 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/772,330
FILING DATE: 23 DECEMBER 1996
ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 2851-A
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644

INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 625 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-996-139-15

Query Match 77.5%; Score 31; DB 3; Length 625;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||:||
Db 21 LLVPLQVTL 29

RESULT 31

US-08-995-659-15
Sequence 15, Application US/08995659
Patent No. 6242213

GENERAL INFORMATION:
APPLICANT: Anderson, Dirk M.
APPLICANT: Galibert, Laurent
APPLICANT: Maraskovsky, Eugene
TITLE OF INVENTION: Ligand for Receptor Activator of NF-kappaB
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation, Law Department
STREET: 51 University Street
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98101

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Apple Power Macintosh
OPERATING SYSTEM: Apple Operating System 7.5.5
SOFTWARE: Microsoft Word for Power Macintosh 6.0.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/995,659
FILING DATE: 22 DECEMBER 1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 60/064,671
FILING DATE: 14 OCTOBER 1997

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/813,509
FILING DATE: 07 MARCH 1997
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: USSN 08/772,330
FILING DATE: 23 DECEMBER 1996

ATTORNEY/AGENT INFORMATION:
NAME: Perkins, Patricia Anne
REGISTRATION NUMBER: 34,693
REFERENCE/DOCKET NUMBER: 2852-A
TELEPHONE: (206)587-0430
TELEFAX: (206)233-0644
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 625 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-995-659-15

Query Match 77.5%; Score 31; DB 4; Length 625;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
|||||:||
Db 21 LLVPLQVTL 29

RESULT 32

US-09-215-649A-15
Sequence 15, Application US/09215649A
Patent No. 6271349

GENERAL INFORMATION:
APPLICANT: Anderson, Dirk M.
Galibert, Laurent
Maraskovsky, Eugene
TITLE OF INVENTION: Receptor Activator of NF-kappaB
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Immunex Corporation, Law Department

```

; STREET: 51 University Street
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Apple Power Macintosh
; OPERATING SYSTEM: Apple Operating System 7.5.5
; SOFTWARE: Microsoft Word for Power Macintosh 6.0.1
;
; CURRENT APPLICATION DATA: US/09/215,649A
; FILING DATE: 17-Dec-1998
; CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/996,139
; FILING DATE: <Unknown>
; APPLICATION NUMBER: USSN 08/813,509
; FILING DATE: 07 MARCH 1997
; APPLICATION NUMBER: USSN 08/772,330
; FILING DATE: 23 DECEMBER 1996
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Perkins, Patricia Anne
; REGISTRATION NUMBER: 34,693
; REFERENCE/DOCKET NUMBER: 2851-A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206)587-0430
; TELEFAX: (206)233-0644
;
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 625 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-09-215-649A-15

Query Match 77.5%; Score 31; DB 4; Length 625;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 21 LLVPLQWTL 29

RESULT 33
US-08-278-635B-6
; Sequence 6, Application US/08278635B
; Patent No. 5683912
;
; GENERAL INFORMATION:
; APPLICANT: ELGOYHEN, ANA BELEN
; APPLICANT: JOHNSON, DAVID S.
; APPLICANT: BOULTER, JAMES R.
; APPLICANT: HEINEMANN, STEPHEN F.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF A NOVEL
; TITLE OF INVENTION: ACETYLCHOLINE-GATED ION CHANNEL RECEPTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GRAY CARY WARE & FREIDENRICH
; STREET: 4365 EXECUTIVE DRIVE, SUITE 1600
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92121
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,635B
; FILING DATE: 06/05/95
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/278,635
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: REITER, STEPHEN E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9989
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-464-258B-6
Query Match 77.5%; Score 31; DB 3; Length 629;
Best Local Similarity 77.8%; Pred. No. 2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

```

; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: REITER, STEPHEN E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9771
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
; US-08-278-635B-6
Query Match 77.5%; Score 31; DB 1; Length 629;
Best Local Similarity 77.8%; Pred. No. 2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 13 LLLPLLLLL 21

RESULT 34
US-08-464-258B-6
; Sequence 6, Application US/08464258B
; Patent No. 6013766
;
; GENERAL INFORMATION:
; APPLICANT: ELGOYHEN, ANA BELEN
; APPLICANT: JOHNSON, DAVID S.
; APPLICANT: BOULTER, JAMES R.
; APPLICANT: HEINEMANN, STEPHEN F.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF A NOVEL
; TITLE OF INVENTION: ACETYLCHOLINE-GATED ION CHANNEL RECEPTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GRAY CARY WARE & FREIDENRICH LLP
; STREET: 4365 EXECUTIVE DRIVE, SUITE 1600
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92121
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/464,258B
; FILING DATE: 06/05/95
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/278,635
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: REITER, STEPHEN E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9989
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
; US-08-464-258B-6
Query Match 77.5%; Score 31; DB 3; Length 629;
Best Local Similarity 77.8%; Pred. No. 2e+02;

```

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPQIILL 9
||||| :||
Db 13 LLLPLLLLL 21

RESULT 35

US-08-471-961-6
; Sequence 6, Application US/08471961
; Patent No. 6100046
; GENERAL INFORMATION:
; APPLICANT: ELGOYHEN, ANA BELEN
; APPLICANT: JOHNSON, DAVID S.
; APPLICANT: BOULTER, JAMES R.
; APPLICANT: HEINEMANN, STEPHEN F.
; TITLE OF INVENTION: CLONING AND EXPRESSION OF A NOVEL
; ACETYLCHOLINE-GATED ION CHANNEL RECEPTOR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GRAY CARY WARE & FREIDENRICH
; STREET: 4365 EXECUTIVE DRIVE, SUITE 1600
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92121
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/471,961
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,635
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: REITER, STEPHEN E.
; REGISTRATION NUMBER: 31,192
; REFERENCE/DOCKET NUMBER: P41 9771
; TELEPHONE: 619-677-1409
; TELEFAX: 619-677-1465
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 629 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-471-961-6

Query Match 77.5%; Score 31; DB 3; Length 629;
Best Local Similarity 77.8%; Pred. No. 2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPQIILL 9
||||| :||
Db 13 LLLPLLLLL 21

RESULT 36

US-09-063-950-2
; Sequence 2, Application US/09063950C
; Patent No. 6225085
; GENERAL INFORMATION:
; APPLICANT: Holtzman, Douglas A.
; TITLE OF INVENTION: NOVEL LRSG PROTEIN AND NUCLEIC ACID MOLECULES AND USES
; FILE REFERENCE: MEI-019
; CURRENT APPLICATION NUMBER: US/09/063,950C
; CURRENT FILING DATE: 1998-04-21

; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 673
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-063-950-2

Query Match 77.5%; Score 31; DB 4; Length 673;
Best Local Similarity 77.8%; Pred. No. 2.1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPQIILL 9
||||| :||
Db 7 LLLPLLLLL 15

RESULT 37

US-08-348-006B-7
; Sequence 7, Application US/08348006B
; Patent No. 5658756
; GENERAL INFORMATION:
; APPLICANT: RODAN, GIDEON A.
; APPLICANT: SCHMIDT, AZRIEL
; APPLICANT: RUTLEDGE, SU JANE
; TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
; TYROSINE PHOSPHATASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: J. MARK HAND
; STREET: 126 E. LINCOLN AVE., P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/348,006B
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/122,032
; FILING DATE: 14-SEP-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: HAND, J., MARK
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 189921A
; TELEPHONE: 908-594-3905
; TELEFAX: 908-594-4720
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-348-006B-7

Query Match 77.5%; Score 31; DB 1; Length 699;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPQIILL 9
||||| :||
Db 52 LLLPLLLLL 60

RESULT 38

US-08-800-825A-7
; Sequence 7, Application US/08800825A
; Patent No. 5866397
; GENERAL INFORMATION:
; APPLICANT: RODAN, GIDEON A.
; APPLICANT: SCHMIDT, AZRIEL
; APPLICANT: RUTLEDGE, SU JANE
; TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: J. MARK HAND - MERCK & CO., INC.
; STREET: 126 E. LINCOLN AVE., P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,825A
; FILING DATE: 14-FEB-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: HAND, J. MARK
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 18992DA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-3905
; TELEFAX: 732-594-4720
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-800-825A-7

Query Match 77.5%; Score 31; DB 2; Length 699;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 52 LLLPLLLLL 60

RESULT 39
US-09-158-657-7
; Sequence 7, Application US/09158657
; Patent No. 6214564
; GENERAL INFORMATION:
; APPLICANT: RODAN, GIDEON A.
; APPLICANT: SCHMIDT, AZRIEL
; APPLICANT: RUTLEDGE, SU JANE
; TITLE OF INVENTION: CDNA ENCODING A NOVEL HUMAN PROTEIN
; TITLE OF INVENTION: TYROSINE PHOSPHATASE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: J. MARK HAND - MERCK & CO., INC.
; STREET: 126 E. LINCOLN AVE., P.O. BOX 2000
; CITY: RAHWAY
; STATE: NEW JERSEY
; COUNTRY: USA
; ZIP: 07065-0900
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/158,657
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/800,825
; FILING DATE: 14-FEB-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: HAND, J. MARK
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 18992DA
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732-594-3905
; TELEFAX: 732-594-4720
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 699 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-158-657-7

Query Match 77.5%; Score 31; DB 4; Length 699;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 52 LLLPLLLLL 60

RESULT 40
US-09-232-200-102
; Sequence 102, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MB
; CURRENT APPLICATION NUMBER: US/09/232,200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 102
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-232-200-102

Query Match 77.5%; Score 31; DB 4; Length 702;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
Db 22 LLLPLLLLL 30

RESULT 41
US-09-232-197-102
; Sequence 102, Application US/09232197A

; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232,197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 102
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-197-102

Query Match 77.5%; Score 31; DB 4; Length 702;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
| | | | | : | |
Db 22 LLLPLLLLL 30

RESULT 42
US-09-232-201-102
; Sequence 102, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MC
; CURRENT APPLICATION NUMBER: US/09/232,201A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 102
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-201-102

Query Match 77.5%; Score 31; DB 4; Length 702;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
| | | | | : | |
Db 22 LLLPLLLLL 30

RESULT 43
US-09-627-376-7

; Sequence 7, Application US/09627376
; Patent No. 6342385
; GENERAL INFORMATION:
; APPLICANT: Qi, Fengxia
; TITLE OF INVENTION: MUTACIN I BIOSYNTHESIS GENES AND PROTEINS
; FILE REFERENCE: UAB-17402/22
; CURRENT APPLICATION NUMBER: US/09/627,376
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 7
; LENGTH: 990
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-627-376-7

Query Match 77.5%; Score 31; DB 4; Length 990;
Best Local Similarity 75.0%; Pred. No. 3.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
| : | | | : | |
Db 357 LVPLQLLL 364

RESULT 44
US-08-658-136-5
; Sequence 5, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, FENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-658-136-5

Query Match 77.5%; Score 31; DB 3; Length 4302;
Best Local Similarity 55.6%; Pred. No. 1.6e+03;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPQLLL 9
:||||:|
Db 2153 VVLPQLVLM 2161

RESULT 45
US-08-460-751-2
: Sequence 2, Application US/08460751
: Patent No. 5891628
: GENERAL INFORMATION:
: APPLICANT: Reeders, Stephen
: APPLICANT: Schneider, Michael
: APPLICANT: Glucksmann, Sandra
: TITLE OF INVENTION: IDENTIFICATION OF POLYCYSTIC KIDNEY
: TITLE OF INVENTION: DISEASE GENE, DIAGNOSTICS AND TREATMENT
: NUMBER OF SEQUENCES: 35
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Pennie & Edmonds
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036-2711
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/460,751
: FILING DATE: 02-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/413,580
: FILING DATE: 03-MAR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Coruzzi, Laura A.
: REGISTRATION NUMBER: 30,742
: REFERENCE/DOCKET NUMBER: 7638-005
: TELEPHONE: (212) 790-9090
: TELEFAX: (212) 869-8864/9741
: TELEX: 66141 PENNIE
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 4303 amino acids
: TYPE: amino acid
: TOPOLOGY: unknown
: MOLECULE TYPE: protein
US-08-460-751-2

Query Match 77.5%; Score 31; DB 2; Length 4303;
Best Local Similarity 55.6%; Pred. No. 1.6e+03;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPQLLL 9
:||||:|
Db 2153 VVLPQLVLM 2161

Search completed: November 6, 2002, 12:06:24
Job time: 11.1111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 23.7778 seconds
(without alignments)
42.042 Million cell updates/sec

Title: US-09-905-083-33
Perfect score: 40
Sequence: 1 LLLPLQILL 9

Scoring table: BLOSUM62
Gapop.10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
1	40	100.0	9	22	AAE08238	Human stratum corn
2	40	100.0	136	22	ABG223378	Novel human diagno
3	40	100.0	253	16	AAE67888	Human stratum corn
4	40	100.0	253	17	AAW05383	Human amyloid prec
5	40	100.0	257	21	AAE21326	Human HSCEE. Homo
6	36	90.0	9	22	AAE08240	Human stratum corn
7	36	90.0	9	22	AAE08241	Human stratum corn
8	35	87.5	61	22	AAO12472	Human polypeptide
9	35	87.5	142	22	AAE63580	Human gastric can
10	35	87.5	156	22	AAE63578	Human gastric can
11	35	87.5	159	22	AAE63582	Human gastric can

12	34	85.0	23	22	ABE43858	Peptide #11364 enc
13	34	85.0	23	22	AAE64837	Human brain expres
14	34	85.0	23	22	AAE37771	Peptide #11808 enc
15	34	85.0	371	16	AAE75642	Bovine conglutinin
16	33	82.5	190	17	AAE98922	Murine APLP1 (aa21
17	33	82.5	653	17	AAE98903	Murine APLP1. Mus
18	33	82.5	770	22	ABG25713	Novel human diagno
19	33	82.5	1404	22	ABG14929	Novel human diagno
20	32	80.0	21	19	AAE48387	Synthetic don-1 po
21	32	80.0	22	19	AAE59850	Amino acid sequenc
22	32	80.0	71	22	AAE89706	Human immune/haema
23	32	80.0	73	22	ABG19347	Novel human diagno
24	32	80.0	147	20	AAE59764	Human normal ovari
25	32	80.0	200	22	AAE02640	Human DNAX surface
26	32	80.0	235	15	AAE49550	Sequence of human
27	32	80.0	235	20	AAE21688	Human CD8 alpha pr
28	32	80.0	247	20	AAE73878	Human chymase prot
29	32	80.0	267	18	AAE22303	Rat CRTI. Rattus
30	32	80.0	301	22	AAE87732	Human T2R02 amino
31	32	80.0	313	22	AAE02639	Human DNAX surface
32	32	80.0	376	21	AAE52588	Secreted modified
33	32	80.0	418	22	AAE92693	Human protein sequ
34	32	80.0	492	13	AAE24863	Sequence of pre-pr
35	32	80.0	590	21	AAE94928	Human secreted pro
36	32	80.0	612	22	AAE94498	Human protein sequ
37	32	80.0	1306	11	AAE04111	Human angiotensin
38	32	80.0	1306	19	AAE68155	Human angiotensin
39	31	77.5	34	22	ABE39600	Peptide #7106 enco
40	31	77.5	34	22	ABE24303	Protein #6302 enco
41	31	77.5	34	22	AAE60304	Human brain expres
42	31	77.5	34	22	AAE72938	Human bone marrow
43	31	77.5	34	22	AAE19781	Peptide #6215 enco
44	31	77.5	34	22	AAE33163	Peptide #7200 enco
45	31	77.5	35	21	AAE56638	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAE08238
ID AAE08238 standard; peptide; 9 AA.
XX
AC AAE08238;
XX
DT 01-NOV-2001 (first entry)
XX
Human stratum corneum chymotrypsin enzyme peptide #3 (residues 5-13).
XX
Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisenese therapy; malignant hyperplasia.
XX
OS Homo sapiens.
XX
PN WO200159158-A1.
XX
PD 16-AUG-2001.
XX
PF 07-FEB-2001; 2001WO-US03977.
XX
PR 11-FEB-2000; 2000US-0502600.
XX
(UYAR-) UNIV ARKANSAS.
XX
O'brien TJ;
XX
WPI; 2001-514676/56.
XX
Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -
XX
PS Claim 25; Page 102; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.

XX Sequence 9 AA;

Query Match 100.0%; Score 40; DB 22; Length 9;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||

DB 1 LLLPLQILL 9

RESULT 2

ABG23378
 ID ABG23378 standard; Protein; 136 AA.

XX

AC ABG23378;

XX

DT 18-FEB-2002 (first entry)

XX

DE Novel human diagnostic protein #23369.

XX

Human; chromosome mapping; gene mapping; gene therapy; forensic;
 food supplement; medical imaging; diagnostic; genetic disorder.

KW

XX

OS Homo sapiens.

XX

PN WO200175067-A2.

XX

PD 11-OCT-2001.

XX

PF 30-MAR-2001; 2001WO-US08631.

XX

PR 31-MAR-2000; 2000US-0540217.

XX

PR 23-AUG-2000; 2000US-0649167.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Drmanac RT, Liu C, Tang YT;

XX

DR WPI; 2001-639362/73.

XX

DR N-PSDB; AAS87565.

XX

New isolated polynucleotide and encoded polypeptides, useful in
 diagnostics, forensics, gene mapping, identification of mutations
 responsible for genetic disorders or other traits and to assess
 biodiversity

XX

PS Claim 20; SEQ ID No 53737; 103pp; English.

XX

The invention relates to isolated polynucleotide (I) and
 polypeptide (II) sequences. (I) is useful as hybridisation probes,
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 and gene mapping, and in recombinant production of (II). The
 polynucleotides are also used in diagnostics as expressed sequence tags
 for identifying expressed genes. (I) is useful in gene therapy techniques
 to restore normal activity of (II) or to treat disease states involving
 (II). (II) is useful for generating antibodies against it, detecting or
 quantitating a polypeptide in tissue, as molecular weight markers and as
 a food supplement. (II) and its binding partners are useful in medical
 imaging of sites expressing (II). (I) and (II) are useful for treating
 disorders involving aberrant protein expression or biological activity.
 The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 136 AA;

Query Match 100.0%; Score 40; DB 22; Length 136;

Best Local Similarity 100.0%; Pred. No. 4.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||

DB 11 LLLPLQILL 19

RESULT 3

AAR67888
 ID AAR67888 standard; Protein; 253 AA.

XX

AC AAR67888;

XX

DT 09-AUG-1995 (first entry)

XX

DE Human stratum corneum chymotrophic recombinant enzyme (SCCE).

XX

KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
 callosities; keratosis pilaris; ichthyoses; eczema.

XX

OS Homo sapiens.

XX

PN WO9500651-A.

XX

PD 05-JAN-1995.

XX

PF 20-JUN-1994; 94WO-IB00166.

XX

PR 18-JUN-1993; 93DK-0000725.

XX

PA (SYMB-) SYMBICOM AB.

XX

PI Egelrud T, Hansson L;

XX

DR WPI; 1995-052088/07.

XX

DR N-PSDB; AAQ81203.

XX

PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme
 - and related vectors, transformed cells and polypeptides, and
 useful for treating skin disorders, e.g. acne or psoriasis, and
 for identification of specific inhibitors.

XX

PS Disclosure; Page 97; 137pp; English.

XX

CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
 CC and skin care products, especially to treat and prevent acne,
 CC keroderma, or other hyperkeratotic conditions (e.g. callosities or
 CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
 CC produced recombinantly following mammal, insect, plant, or
 CC microorganism transformation with plasmid pS507.

XX Sequence 253 AA;

Query Match 100.0%; Score 40; DB 16; Length 253;

Best Local Similarity 100.0%; Pred. No. 7.8;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

|||||

DB 5 LLLPLQILL 13

```

RESULT 4
AAW05383
ID AAW05383 standard; Protein; 253 AA.
XX
AC AAW05383;
XX
DT 31-DEC-1996 (first entry)
XX
DE Human amyloid precursor protein protease.
XX
KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;
KW therapy.
OS Homo sapiens.
XX
XX WO9631122-A1.
PN
PD 10-OCT-1996.
XX
PF 02-APR-1996; 96WO-US04294.
XX
PR 04-APR-1995; 95US-0416257.
XX
PA (ELIL ) LILLY & CO ELI.
XX
PI Dixon EP, Johnstone EM, Little SP;
XX
XX WPI; 1996-464694/46.
DR
N-PSDB; AAT39783.
XX
XX New isolated human amyloid precursor protein protease - used to
PT develop prods. for the treatment or diagnosis of associated
PT conditions, esp. Alzheimer's disease
XX
PS Claim 1; Page 44-45; 55pp; English.
XX
XX Human amyloid precursor protein protease (AAW05383) is involved in
CC the processing or clearance of amyloid precursor protein to form
CC beta-amyloid peptide. Its amino acid sequence was deduced from
CC a cDNA clone (AAT39783) obtd. from a human lung library. Recombinant
CC protease can be produced in transformed or transfected prokaryotic
CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
CC used to develop products for the design and testing of cnds. useful
CC for treating or preventing conditions associated with beta-amyloid
CC peptide, esp. Alzheimer's disease.
XX
SQ Sequence 253 AA;
XX
Query Match 100.0%; Score 40; DB 17; Length 253;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLLPLQLLL 9
DB 5 LLLPLQLLL 13
IIIIIIII
RESULT 5
AAB21326
ID AAB21326 standard; Protein; 257 AA.
XX
AC AAB21326;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human HSCEE.
XX
XX Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCEE;
KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;
KW serine protease; cytostatic; cancer; prostrate cancer.
XX

```

```

OS Homo sapiens.
XX
XX WO200053776-A2.
XX
PD 14-SEP-2000.
XX
XX 09-MAR-2000; 2000WO-CA00258.
XX
XX 11-MAR-1999; 99US-0124260.
PR 01-APR-1999; 99US-0127386.
PR 21-JUL-1999; 99US-0144919.
XX
XX (MOUN ) MOUNT SINAI HOSPITAL.
PA
XX Yousef GM, Diamandis EP;
XX
XX WPI; 2000-587440/55.
DR
XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
PT protein mediated disorders, especially cancer.
XX
XX Example 4; Fig 17; 184pp; English.
XX
XX The present sequence is human stratum corneum chymotryptic enzyme
CC (HSCEE), a member of the kallikrein multi-gene family. Kallikreins and
CC kallikrein-like proteins are a subgroup of the serine protease enzyme
CC family. They catalyse the selective cleavage of specific polypeptide
CC precursors to release peptides with potent biological activity. Nucleic
CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
CC treatment, monitoring and diagnosis of cancers, especially prostate
CC cancer. They can also be used to identify a substance that can associate
CC with or mediate the biological activity of the proteins. Antibodies can
CC be used to treat conditions mediated by the kallikrein-like proteins.
XX
SQ Sequence 257 AA;
XX
Query Match 100.0%; Score 40; DB 21; Length 257;
Best Local Similarity 100.0%; Pred. No. 8;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLLPLQLLL 9
DB 5 LLLPLQLLL 13
IIIIIIII
RESULT 6
AAE08240
ID AAE08240 standard; peptide; 9 AA.
XX
XX AAE08240;
AC
XX 01-NOV-2001 (first entry)
DT
XX Human stratum corneum chymotrypsin enzyme peptide #5 (residues 6-14).
XX
XX Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.
XX
XX Homo sapiens.
OS
XX WO200159158-A1.
PN
XX 16-AUG-2001.
PD
XX 07-FEB-2001; 2001WO-US03977.
XX
XX 11-FEB-2000; 2000US-0502600.
PR
XX (UYAR-) UNIV ARKANSAS.
PA
XX O'brien TJ;
PI

```



```

Db      4 ILLPLQILV 12
      :|||||:
RESULT 9
AAB63580
ID AAB63580 standard; Protein; 142 AA.
XX
AC AAB63580;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human gastric cancer associated antigen protein sequence SEQ ID NO:942.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
XX
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;
XX
PN WI; 2001-025274/03.
XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
PS Example 1; Page 624; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX
SQ Sequence 156 AA;
XX
PS Query Match 87.5%; Score 35; DB 22; Length 156;
XX Best Local Similarity 77.8%; Pred. No. 41;
XX Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLLPLQILL 9
Db 47 VLLPLQLLL 55
      :|||||:
RESULT 11
AAB63582
ID AAB63582 standard; Protein; 159 AA.
XX
AC AAB63582;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human gastric cancer associated antigen protein sequence SEQ ID NO:944.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
XX
PR 10-SEP-1999; 99US-0153454.
XX

```

```

Db      4 ILLPLQILV 12
      :|||||:
RESULT 9
AAB63580
ID AAB63580 standard; Protein; 142 AA.
XX
AC AAB63580;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human gastric cancer associated antigen protein sequence SEQ ID NO:942.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
XX
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;
XX
PN WI; 2001-025274/03.
XX
PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
PS Example 1; Page 625; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX
SQ Sequence 142 AA;
XX
PS Query Match 87.5%; Score 35; DB 22; Length 142;
XX Best Local Similarity 77.8%; Pred. No. 37;
XX Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLLPLQILL 9
Db 47 VLLPLQLLL 55
      :|||||:
RESULT 10
AAB63578
ID AAB63578 standard; Protein; 156 AA.
XX
AC AAB63578;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human gastric cancer associated antigen protein sequence SEQ ID NO:940.
XX

```

PA (LUDW-) LUDWIG INST CANCER RES.

XX Obata Y;

XX WPI; 2001-025274/03.

XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -

XX Example 1; Page 625-626; 799pp; English.

XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP), respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.

XX Sequence 159 AA;

Query Match 87.5%; Score 35; DB 22; Length 159;
Best Local Similarity 77.8%; Pred. No. 41;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLLPLQLLL 9
Db 47 VLLPLQLLL 55
:||||:|

RESULT 12

ABB43858
ID ABB43858 standard; Peptide; 23 AA.

XX ABB43858;

XX 04-FEB-2002 (first entry)

XX Peptide #11364 encoded by human foetal liver single exon probe.

XX Human; foetal liver; gene expression; single exon nucleic acid probe.

XX Homo sapiens.

XX WO200157277-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00669.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483447/52.

XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human foetal liver -

PS Claim 27; SEQ ID NO 36493; 639pp + sequence listing; English.

XX

CC The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 23 AA;

Query Match 85.0%; Score 34; DB 22; Length 23;

Best Local Similarity 87.5%; Pred. No. 9.2;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQLLL 9
Db 11 LLLPLQLLL 18
:||||:|

RESULT 13

AAM64837
ID AAM64837 standard; Protein; 23 AA.

XX AAM64837;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 36942.

XX Human; brain expressed exon; gene expression analysis; probe;

XX microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

XX epilepsy; cancer.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00667.

XX 04-FEB-2000; 2000US-0180312.

XX 26-MAY-2000; 2000US-0207456.

XX 30-JUN-2000; 2000US-0608408.

XX 03-AUG-2000; 2000US-0632366.

XX 21-SEP-2000; 2000US-0234687.

XX 27-SEP-2000; 2000US-0236359.

XX 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human
XX brains -

XX Example 4; SEQ ID NO: 36942; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid
CC probes which are derived from genomic sequences expressed in the human
CC brain. They can be used to measure gene expression in brain cell samples,
CC which may enable the diagnosis and improved treatment of nervous system
CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
CC epilepsy and cancers. The present sequence is a protein encoded by one of
CC the probes of the invention.

XX Sequence 23 AA;

Query Match 85.0%; Score 34; DB 22; Length 23;
Best Local Similarity 87.5%; Pred. No. 9.2;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQILL 9
|||||:||
Db 11 LLLPLQLLL 18

RESULT 14
AAR37771
ID AAR37771 standard; Protein; 23 AA.
XX AC AAR37771;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #11808 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00663.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human placenta -
XX Claim 27; SEQ ID No 38040; 654pp; English.

The present invention relates to single exon nucleic acid probes (SENP;
see AAT31315-AA157546). The present sequence is a peptide encoded by one
CC such probe. The probes are useful for producing a microarray for
CC predicting, measuring and displaying gene expression in samples derived
CC from human placenta. The probes are useful for antenatal diagnosis of
CC human genetic disorders.
XX Sequence 23 AA;
Query Match 85.0%; Score 34; DB 22; Length 23;
Best Local Similarity 87.5%; Pred. No. 9.2;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 LLLPLQILL 9
|||||:||
Db 11 LLLPLQLLL 18
RESULT 15
AAR75642
ID AAR75642 standard; Protein; 371 AA.
XX AC AAR75642;
XX

DT 10-FEB-1996 (first entry)
XX Bovine conglutinin.
DE Conglutinin; therapeutic; diagnostic.
XX Bos taurus.
XX Key Location/Qualifiers
FH Peptide 1..20
FT /note= "signal peptide"
FT Protein 21..371
FT /note= "mature bovine conglutinin"
FT Region 270..371
FT /note= "carbohydrate-binding region"
FT Region 217..371
FT /note= "carbohydrate-binding region"
XX WO9516697-A1.
PN 22-JUN-1995.
PD 14-DEC-1994; 94WO-US14656.
XX 16-DEC-1993; 93US-0168458.
XX (IMMU-) APPLIED IMMUNE SCI INC.
PA (UYBO-) UNIV BOSTON.
XX Lee YM, Leiby KR, Okarma TB, Sastry K;
XX WPI; 1995-231510/30.
DR N-PSDB; AAQ92554.
XX New nucleic acid encoding bovine conglutinin and its fragments - and
PT related vectors, host cell, etc. useful e.g. in modulating
PT conglutinin expression or for selective removal of cpds., carrying
PT particular carbohydrate residues.
XX Claim 1; Fig 1A-1B; 29pp; English.
XX Conglutinin, including the complete sequence, the mature protein or
CC the carbohydrate-binding fragments, are used to bind cells carrying
CC the C1q receptor. Solid supports are used to remove compounds
CC (especially proteins or peptides, e.g. immune complexes) having a
CC carbohydrate component able to bind to the lectin region of
CC conglutinin, optionally with recovery of bound compounds.
XX Sequence 371 AA;
Query Match 85.0%; Score 34; DB 16; Length 371;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLLPLQILL 9
|||||:||
Db 2 LLLPLSVLL 10
RESULT 16
AAR98922
ID AAR98922 standard; Protein; 190 AA.
XX AC AAR98922;
XX DT 29-OCT-1996 (first entry)
XX Murine APLP1 (aa21-210).
XX APLP1; amyloid precursor-like protein 1; APP;
KW amyloid beta-protein precursor; promoter; transcription;
KW upstream stimulatory factor; Alzheimer's disease.
XX

```

OS Mus sp.
PN WO9615265-A1.
XX
XX 23-MAY-1996.
XX
XX 09-NOV-1995; 95WO-US14416.
XX
XX 10-NOV-1994; 94US-0339152.
XX
XX (GEOH) GEN HOSPITAL CORP.
XX
XX Kovacs DM, Tanzi RE;
XX
XX WPI; 1996-259860/26.
XX
XX Modulating transcription from amyloid beta-protein precursor
PT promoter - using upstream stimulatory factor (USF), and USF-binding
PT compounds such as amyloid precursor-like proteins APLP1 and APLP2
XX
XX Example 1; Page 78; 130pp; English.
XX
XX 3 Regions (AAR98922-24) of the mouse amyloid precursor-like
CC protein APLP1 (see also AAR98903) show a high degree of homology to
CC the human amyloid precursor protein (APP). The 3 regions
CC respectively correspond to amino acids 21-210, 316-470 and 609-654
CC in extracellular domain I, extracellular domain II and the
CC cytoplasmic domain of APLP1. APLP1 is an upstream stimulatory
CC factor binding protein capable of down-regulating expression
CC from the APP gene promoter. It is a new member of the APP-like
CC family.
XX
XX Sequence 190 AA;
SQ
Query Match 82.5%; Score 33; DB 17; Length 190;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 1 LLLPLSLLL 9

RESULT 17
AAR98903
ID AAR98903 standard; Protein: 653 AA.
XX
XX AAR98903;
XX
XX 29-OCT-1996 (first entry)
XX
XX Murine APLP1.
XX
KW APLP1; amyloid precursor-like protein 1; APP;
KW amyloid beta-protein precursor; promoter; transcription;
KW upstream stimulatory factor; Alzheimer's disease.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
FH Peptide 1..20
FT /label= Sig_peptide
FT Modified-site 464..466
FT /label= N-glycosylation_site
FT Modified-site 554..556
FT /label= N-glycosylation_site
FT Domain 584..606
FT /label= Membrane-spanning_domain
FT Peptide 631..639
FT /note= "peptide used to raise antisera"
FT Modified-site 642..647
FT /note= "potential tyrosine phosphorylation site"
XX

```

```

PN WO9615265-A1.
XX
XX 23-MAY-1996.
XX
XX 09-NOV-1995; 95WO-US14416.
XX
XX 10-NOV-1994; 94US-0339152.
XX
XX (GEOH) GEN HOSPITAL CORP.
XX
XX Kovacs DM, Tanzi RE;
XX
XX WPI; 1996-259860/26.
XX
XX N-PSDB; AAT34694.
XX
XX Modulating transcription from amyloid beta-protein precursor
PT promoter - using upstream stimulatory factor (USF), and USF-binding
PT compounds such as amyloid precursor-like proteins APLP1 and APLP2
XX
XX Example 1; Page 70-72; 130pp; English.
XX
XX Murine amyloid precursor-like protein 1 (APLP1) (AAR98903) is a novel
CC member of the amyloid beta-protein precursor (APP)-like family,
CC and includes regions (see also AAR98922-24) of homology to APP. It
CC acts as an upstream stimulatory factor binding cpd., and is capable
CC of down-regulating transcription from the APP gene promoter. Its
CC amino acid sequence was deduced from a cDNA clone (AAT34694) obtd.
CC from a mouse brain cDNA library. Elucidation of factors involved
CC in the transcriptional regulation of the APP gene will help
CC determine the mechanism which results in the formation of amyloid
CC deposits and causes Alzheimer's disease.
XX
XX Sequence 653 AA;
SQ
Query Match 82.5%; Score 33; DB 17; Length 653;
Best Local Similarity 77.8%; Pred. No. 4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 21 LLLPLSLLL 29

RESULT 18
ABG25713
ID ABG25713 standard; Protein: 770 AA.
XX
XX ABG25713;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #25704.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
XX
XX WO200175067-A2.
XX
XX 11-OCT-2001.
XX
XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
XX
XX 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
XX
XX N-PSDB; AAS89900.
XX

```


XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 56072; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 770 AA;
Query Match 82.5%; Score 33; DB 22; Length 770;
Best Local Similarity 77.8%; Pred. No. 4.7e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLLPLQLLL 9
II IIIII;
DB 327 LLAPLQILM 335
RESULT 19
ABG14929
ID ABG14929 standard; Protein; 1404 AA.
XX
AC ABG14929;
XX
XX 18-FEB-2002 (first entry)
XX
XX Novel human diagnostic protein #14920.
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
XX Homo sapiens.
OS
XX WO200175067-A2.
PN
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX 31-MAR-2000; 2000US-0540217.
PR
XX 23-AUG-2000; 2000US-0649167.
PR
XX (HYSE-) HYSEQ INC.
PA
XX Drmanac RT, Liu C, Tang YT;
XX
XX WPI: 2001-639362/73.
DR
XX N-PSDB; AAS79116.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
XX Claim 20; SEQ ID No 45288; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1404 AA;
Query Match 82.5%; Score 33; DB 22; Length 1404;
Best Local Similarity 77.8%; Pred. No. 8.6e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLLPLQLLL 9
II IIIII :II
DB 191 LLLPLDLLL 199
RESULT 20
AAW48387
ID AAW48387 standard; Protein; 21 AA.
XX
AC AAW48387;
XX
XX 17-AUG-1998 (first entry)
XX
XX Synthetic don-1 polypeptide EGF domain leader sequence.
DE
XX
XX Murine; don-1 gene; melanoma; treatment; adenocarcinoma; EGF domain;
KW epithelial cell; proliferation; stimulation; treatment; tumours;
KW skin; oesophagus; lung; breast; liver; pancreas; colon; prostate;
KW gastrointestinal tract; uterus; wound healing; transmembrane;
KW leader sequence.
XX
XX Synthetic.
XX
XX WO9807736-A1.
PN
XX 26-FEB-1998.
PD
XX
XX 18-AUG-1997; 97WO-US14585.
PF
XX 19-NOV-1996; 96US-0753007.
PR
XX 19-AUG-1996; 96US-0699591.
PR
XX (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
PA
XX Busfield SJ, Gearing DP;
XX
XX WPI: 1998-169084/15.
DR
XX Mouse and human don-1 polypeptide(s) - useful for treatment of

PT melanomas and adenocarcinoma(s), and for wound healing

XX
PS Example 2; Page 91; 121pp; English.

XX The sequence is that of a leader sequence for the EGF domain of
CC human don-1. It was used as part of the characterisation of don-1.
CC Don-1 polypeptides stimulate proliferation of epithelial cells
CC and thus are implicated in melanomas and adenocarcinomas in which
CC epithelial cells proliferate out of control. Compounds that
CC interfere with don-1 mediated cell proliferation can be used
CC in the treatment of tumours such as melanomas and adenocarcinomas
CC of the skin, oesophagus, lung, breast, liver, pancreas,
CC gastrointestinal tract, colon, prostate or uterus. Alternatively,
CC don-1 polypeptides can be used to stimulate epithelial cell
CC proliferation, e.g. for wound healing.

XX Sequence 21 AA;

Query Match 80.0%; Score 32; DB 19; Length 21;
Best Local Similarity 77.8%; Pred. No. 20;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 8 LLLPLALL 16

RESULT 21

AAW59850
ID AAW59850 standard; peptide; 22 AA.

XX
AC AAW59850;

XX 26-OCT-1998 (first entry)

XX Amino acid sequence of mammalian T8 signal sequence.

XX Human placental alkaline phosphatase; anchoring sequence; agonist;
KW antagonist; cellular receptor.

XX Mammalia.

XX WO9825146-A1.

XX 11-JUN-1998.

XX 01-DEC-1997; 97WO-GB03303.

XX 03-DEC-1996; 96US-0758307.

XX (GLAXO) GLAXO GROUP LTD.

XX Dower WJ, Gates CM, Heinkel GL, Lalonde G, Mattheakis LC;
PI Paddon CJ, Schatz PJ;

XX WPI; 1998-377243/32.

XX Screening libraries of complexes for compounds with desired
PT properties - using tethers which may be modified by a reporter
PT molecule, modification indicates that a complex with the tether has
PT the desired property

XX Disclosure; Page 68; 119pp; English.

XX This is the amino acid sequence of the mammalian T8 signal sequence
CC used in the method of the invention. The methods are useful for
CC screening libraries of complexes for compounds with desired properties,
CC especially the capacity to agonise, bind to or antagonise cellular
CC receptors. The processes allow distinctions to be made between
CC compounds that merely bind to a receptor and compounds that are
CC capable of transducing biological signals through the receptor.

XX Sequence 22 AA;

Query Match 80.0%; Score 32; DB 19; Length 22;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
DB 8 LLLPLALL 16

RESULT 22

AAM89706
ID AAM89706 standard; Protein; 71 AA.

XX
AC AAM89706;

XX 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen SEQ ID NO:17299.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.

XX Homo sapiens.

XX WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 14-AUG-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 14-AUG-2000; 2000US-0225759.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX 22-AUG-2000; 2000US-0226868.

XX 22-AUG-2000; 2000US-0227182.

XX 23-AUG-2000; 2000US-0227009.

XX 30-AUG-2000; 2000US-0228924.

XX 01-SEP-2000; 2000US-0229287.

XX 01-SEP-2000; 2000US-0229343.

XX 01-SEP-2000; 2000US-0229344.

XX 01-SEP-2000; 2000US-0229345.

XX 05-SEP-2000; 2000US-0229509.

XX 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.

PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-483426/52.
DR N-PSDB; AAK62487.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis.
XX
PS Claim 11; SEQ ID NO 17299; 3071pp + Sequence Listing; English.
XX
CC AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/haematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/haematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
SQ Sequence 71 AA;
Query Match 80.0%; Score 32; DB 22; Length 71;
Best Local Similarity 75.0%; Pred. No. 66;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLLPLQL 8
DB 14 ILIPLQL 21
RESULT 23
ABG19347
ID ABG19347 standard; Protein; 73 AA.
XX
AC ABG19347;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #19338.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder..
 XX Homo sapiens.
 OS
 XX WO200175067-A2.
 PN
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI; 2001-639362/73.
 DR N-PSDB; AAS83534.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX
 PS Claim 20; SEQ ID No 49706; 103pp; English.
 XX
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 73 AA;
 Query Match 80.0%; Score 32; DB 22; Length 73;
 Best Local Similarity 75.0%; Pred. No. 68;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLLPLQL 8
 Db 42 LLLPLQLM 49
 RESULT 24
 AAY59764
 ID AAY59764 standard; Protein; 147 AA.
 XX
 AC AAY59764;
 XX
 DT 18-JAN-2000 (first entry)
 XX
 DE Human normal ovarian tissue derived protein 41.
 KW Human; ovary; screening; ovarian cancer; treatment.

XX Homo sapiens.
 OS
 XX DE19816395-A1.
 PN
 XX 07-OCT-1999.
 PD
 XX 03-APR-1998; 98DE-1016395.
 PF
 XX 03-APR-1998; 98DE-1016395.
 PR
 XX (META-) METAGEN GES GENOMFORSCHUNG MBH.
 PA
 XX Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pillarsky C, Dahl E;
 PI
 XX WPI; 1999-552352/47.
 DR N-PSDB; AAZ41240.
 DR
 XX Nucleic acid sequences potentially useful in diagnosis or therapy of
 PT ovarian cancer -
 PT
 XX
 PS Claim 23; Page 232; 274ppp; German.
 XX
 CC This invention describes novel nucleic acid sequences that are highly
 CC expressed in normal ovary tissue. Artificial chromosomes and cosmid
 CC clones containing the sequences can be used as gene transfer vehicles.
 CC The sequences can be used to produce DNA fragments containing
 CC full-length genes. Host cells transformed with the sequences can be used
 CC to produce polypeptides or polypeptide fragments, which can be used to
 CC screen phage displays for polypeptides that bind to them, or as tools for
 CC identifying agents active against ovarian cancer, or to prepare
 CC medicaments for treating ovarian cancer. The cDNA sequences can be used
 CC to obtain genomic genes, their promoters, enhancers, silencers, exon
 CC structures, intron structures and their splice variants. AAY59724-Y59837
 CC represent protein fragments encoded by the cDNA sequences represented in
 CC AAZ41222-441324 which are derived from normal human ovarian tissue.
 XX
 SQ Sequence 147 AA;
 Query Match 80.0%; Score 32; DB 20; Length 147;
 Best Local Similarity 77.8%; Pred. No. 1.4e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPLQL 9
 Db 42 LLLPLLLL 50
 RESULT 25
 AAE02640
 ID AAE02640 standard; Protein; 200 AA.
 XX
 AC AAE02640;
 XX
 DT 06-AUG-2001 (first entry)
 XX
 DE Human DNAX surface short form protein (DSP-1S).
 XX
 KW Human; DNAX surface protein; DSP-1; forensic science; therapy;
 KW abnormal physiology; allergic condition; asthma; cancer;
 KW autoimmune disease; diabetes mellitus; drug screening.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT /label= Signal_peptide
 FT Protein 22..200
 FT /note= "Mature DNAX surface short form protein (DSP-1L)."
 XX
 PN WO200136463-A2.
 XX
 PD 25-MAY-2001.

XX 15-NOV-2000; 2000WO-US31167.
 XX
 PR 15-NOV-1999; 99US-0439735.
 XX
 PA (SCHE) SCHERING CORP.
 PI Zlot CH, Adema GJ, Figdor C, Phillips JH;
 XX
 DR WPI: 2001-367562/38.
 DR N-PSDB; AAD06866.
 XX
 XX New mammalian proteins designated dendritic cell specific transmembrane
 PT protein and DNAX surface protein and the nucleic acids encoding the
 PT polypeptides -
 XX
 XX Claim 2: Page 17; 96pp; English.
 XX
 CC The present sequence is human DNAX surface short form protein (DSP-1S).
 CC The DSP-1S is useful in forensic sciences, e.g. to distinguish
 CC rodent from human or as a marker to distinguish between different
 CC cells exhibiting differential expression or modification patterns.
 CC The DSP-1S and its antibodies are used for preparing kits for use in
 CC molecular biology, immunology or physiology, and in treatment of
 CC conditions associated with abnormal physiology or development, e.g.
 CC allergic conditions (asthma), cancer and autoimmune diseases (diabetes
 CC mellitus). Drug screening using DSP-1S or its fragments can be
 CC performed to identify compounds having binding affinity to or other
 CC relevant isolation effects on the function of DSP-1S. The DSP-1S can
 CC also be used in diagnostic kits and methods for detecting the presence
 CC of another DSP-1S or binding partner.
 XX
 XX Sequence 200 AA;
 SQ
 Query Match 80.0%; Score 32; DB 22; Length 200;
 Best Local Similarity 77.8%; Pred. NO. 1.9e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPLQILL 9
 Db 8 LLLPLALLL 16
 ||||| :||
 RESULT 26
 AAR49550
 ID AAR49550 standard; Protein; 235 AA.
 XX
 AC AAR49550;
 XX
 DT 20-AUG-1994 (first entry)
 XX
 DE Sequence of human CD8.
 XX
 KW gp39; T-cell antigen; CD40 ligand; B-cell proliferation;
 KW CD8; fusion protein; CD8; fusion protein.
 XX
 OS Homo sapiens.
 XX
 PN EP585943-A.
 XX
 PD 09-MAR-1994.
 XX
 PF 03-SEP-1993; 93EP-0114153.
 XX
 PR 04-SEP-1992; 92US-0940605.
 XX
 XX (BRIM) BRISTOL-MYERS SQUIBB CO.
 PA
 PI Aruffo AA, Hollenbaugh D, Ledbetter JA, Aruffo A;
 XX
 DR WPI: 1994-076264/10.
 DR N-PSDB; AAR495986.
 XX

PT New nucleic acid encoding human gp39 T cell antigen - which is a
 PT ligand for the CD40 receptor, causing proliferation and
 PT differentiation of B cells and some cancer cells
 XX
 PS Disclosure; Fig 9; 39pp; English.
 XX
 CC The complete nucleic acid sequence of human gp39 (hgp39) protein
 CC (corresp. to cDNA) and the complete AA sequence of hgp39 are
 CC presented in AAQ57984 and AAR45548 respectively and contd. in plasmid
 CC CDM8-hgp39, deposited with the ATCC as E. coli, CDM8 WC1061/p3-hgp39
 CC and assigned accession No. 69050. The human T cell antigen gp39 is a
 CC ligand for the CD40 receptor. Soluble gp39 may be produced using the
 CC expression vector CD8-gp39. Chimeric genes may be constructed by
 CC fusing sequences encoding the extracellular domains of gp39 and
 CC CD8, pref. murine or human CD8 protein. Plasmid p3-shgp39 encoding
 CC the fusion protein of the extracellular domains from gp39 and CD8
 CC is deposited in E.coli as ATCC 69049.
 XX
 XX Sequence 235 AA;
 SQ
 Query Match 80.0%; Score 32; DB 15; Length 235;
 Best Local Similarity 77.8%; Pred. NO. 2.2e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPLQILL 9
 Db 8 LLLPLALLL 16
 ||||| :||
 RESULT 27
 AAY21688
 ID AAY21688 standard; Protein; 235 AA.
 XX
 AC AAY21688;
 XX
 DT 18-AUG-1999 (first entry)
 XX
 DE Human CD8 alpha protein.
 XX
 KW T lymphocyte inhibition; CD8; immunosuppressive therapy; transplantation;
 KW autoimmune disease; allergy; asthma; viral infection; cytotoxic;
 KW corticosteroid; human; CD8 alpha protein.
 XX
 OS Homo sapiens.
 XX
 PN WO9921576-A1.
 XX
 PD 06-MAY-1999.
 XX
 PF 28-OCT-1998; 98WO-GB03235.
 XX
 PR 28-OCT-1997; 97GB-0022779.
 XX
 XX (ISIS-) ISIS INNOVATION LTD.
 PA
 PI Gao GF, Gerth UC, Jakobsen BK, Sewell AK;
 XX
 DR WPI: 1999-385058/32.
 DR N-PSDB; AAX80967.
 XX
 PT Inhibiting activity of T cells against target cells useful for
 PT treating autoimmune diseases and allergy
 XX
 PS Disclosure; Fig 1A; 79pp; English.
 XX
 CC The invention describes a method for inhibiting activity of T lymphocytes
 CC against a target cell by treating the cell with a soluble form of a CD8
 CC molecule. The method is used as immunosuppressive therapy, e.g. in
 CC patients undergoing transplantation, but also for treating autoimmune
 CC diseases and allergy, e.g. exacerbation of asthma caused by viral
 CC infection. The CD8 protein inhibits cytotoxic T cells in vitro or in
 CC vivo, amnd its inhibitory activity can be adjusted by mutation. Treatment
 CC with CD8 protein provides more selective immunosuppression than use of

CC corticosteroids. The present sequence represents a human CD8 alpha protein.

XX
SQ Sequence 235 AA;

Query Match 80.0%; Score 32; DB 20; Length 235;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9
||||| :||
Db 8 LLLPLALL 16

RESULT 28

AAW73878
ID AAW73878 standard; Protein; 247 AA.

XX AC AAW73878;

XX DT 31-MAR-1999 (first entry)

XX DE Human chymase protein sequence.

XX KW Chymase; human; secretion expression vector; protein production.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..21

FT Protein /note= "signal peptide"

FT Protein 22..247

FT Protein /note= "mature chymase"

XX PN JP10327870-A.

XX PD 15-DEC-1998.

XX PF 03-JUN-1997; 97JP-0145186.

XX PR 03-JUN-1997; 97JP-0145186.

XX PA (GREC) GREEN CROSS CORP.

XX DR WPI: 1999-135568/12.

XX DR N-PSDB; AAX01231.

XX PT Recombinant human chymase - used to prepare the human chymase protein

XX PS Claim 2; Page 12-13; 16pp; Japanese.

XX CC This sequence is the human chymase protein. The invention relates to a method for the preparation of active human chymase by transformation of a yeast host with a secretion expression vector containing DNA encoding mature human chymase, and culturing of the cells to give active human chymase. The recombinant DNA can be used to produce the human chymase protein. Using the method, human chymase can be produced easily and cost effectively.

XX SQ Sequence 247 AA;

Query Match 80.0%; Score 32; DB 20; Length 247;
Best Local Similarity 77.8%; Pred. No. 2.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9
||||| :||
Db 2 LLLPLLLL 10

RESULT 29

AAW22303

ID AAW22303 standard; Protein; 267 AA.

XX AC AAW22303;

XX DT 24-NOV-1997 (first entry)

XX DE Rat CRTI.

XX KW Rat; CRTI; expression; liver cancer; tissue; antibody; probe; detection; assay; diagnosis.

XX OS Rattus rattus.

XX PN WO9710333-A1.

XX PD 20-MAR-1997.

XX PF 17-SEP-1996; 96WO-JP02654.

XX PR 30-AUG-1996; 96JP-0229469.

XX PR 14-SEP-1995; 95JP-0236264.

XX PR 27-NOV-1995; 95JP-0331023.

XX PR 20-JUN-1996; 96JP-0179885.

XX PA (SUME) SUMITOMO ELECTRIC IND CO.

XX PI Kishimoto T, Kokura K, Kumagai Y, Makino Y, Tamura T;

XX DR WPI: 1997-202229/18.

XX DR N-PSDB; AAT73337.

XX PT Proteins with elevated expression in liver cancer, and related DNA - for production of antibodies useful in the diagnosis and monitoring of liver cancer

XX PS Claim 1; Pages 69-70; 119pp; Japanese.

XX CC The present sequence is rat CRTI, which has an elevated expression in liver cancer tissue. Anti-CRTI antibodies or CRTI cDNA can be used to detect or assay for CRTI or CRTI mRNA, indicating in which tissues CRTI expression is elevated, and therefore allowing liver cancer to be diagnosed and monitored.

XX SQ Sequence 267 AA;

Query Match 80.0%; Score 32; DB 18; Length 267;
Best Local Similarity 100.0%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPLQLL 8

Db 188 LLPLQLL 194

RESULT 30

AAW87732

ID AAW87732 standard; Protein; 301 AA.

XX AC AAW87732;

XX DT 16-MAY-2001 (first entry)

XX DE Human T2R02 amino acid sequence SEQ ID NO:3.

XX KW Human; rat; mouse; T2R; taste receptor; G-protein coupled receptor; taste transduction G-protein coupled receptor; identification; tongue; taste sensory neuron; taste cell; taste modulator; food; taste signalling pathway.

XX OS Homo sapiens.

XX PN WO200118050-A2.

PD 15-MAR-2001.
 XX
 XX 08-SEP-2000; 2000WO-US24821.
 XX
 XX 10-SEP-1999; 99US-0393634.
 PR 22-FEB-2000; 2000US-0510332.
 XX
 XX (REGC) UNIV CALIFORNIA.
 PA (USSH) US DEPT HEALTH & HUMAN SERVICES.
 XX
 XX Zuker CS, Adler JE, Ryba N, Mueller K, Hoon M;
 PI WPI; 2001-211396/21.
 XX N-PSDB; AAF92503.
 DR
 XX Nucleic acids encoding the T2R family of G-protein coupled taste
 PT receptors, useful for identifying taste modulators that can be used in
 PT food and pharmaceutical industries to customize taste, for e.g. to
 PT decrease the bitter taste of food -
 XX
 XX Claim 19; Page 161; 249pp; English.
 PS
 XX AAF92502 to AAF92572 represent nucleic acids which encode taste
 CC transduction G-protein coupled receptors designated T2R proteins.
 CC AAB87731 to AAB87824 represent T2R proteins, and AAB87825 to AAB87830
 CC represent T2R family consensus sequences from the present invention.
 CC The T2R proteins are taste modulators. The nucleic acids are useful as
 CC probes for the identification of taste cells, as the nucleic acids are
 CC specifically expressed in taste cells. They also serve as tools for the
 CC generation of taste topographic maps that elucidate the relationship
 CC between the taste cells of the tongue and taste sensory neurons leading
 CC to taste centres in the brain. The taste modulators are useful for
 CC pharmacological and genetic modulation of taste signalling pathways.
 CC Modulatory compounds comprising T2R proteins can therefore be used in
 CC food and pharmaceutical industries to customise taste, for e.g. to
 CC decrease the bitter taste of food or drugs.
 XX
 XX Sequence 301 AA;
 SQ
 Query Match 80.0%; Score 32; DB 22; Length 301;
 Best Local Similarity 75.0%; Pred. No. 2.8e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LLLPQLLL 9
 Db 42 LMPQLLL 49
 1:1:1111
 1:1:1111

RESULT 31
 AAE02639
 ID AAE02639 standard; Protein; 313 AA.
 XX
 XX AAE02639;
 AC
 XX 06-AUG-2001 (first entry)
 DT
 XX Human DNAX surface long form protein (DSP-1L).
 DE
 XX Human; DNAX surface protein; DSP-1; forensic science; therapy;
 KW abnormal physiology; allergic condition; asthma; cancer;
 KW autoimmune disease; diabetes mellitus; drug screening.
 XX
 XX Homo sapiens.
 OS
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /label= Signal_peptide
 FT Protein 22..313
 FT /note= "Mature DNAX surface long form protein (DSP-1L)"
 FT 193..209
 FT /label= Transmembrane_segment
 FT 243..248
 FT /label= ITIM_motif
 FT

FT Domain 265..272
 FT /label= ITIM_motif
 FT 279..284
 FT /label= ITIM_motif
 XX
 XX WO200136463-A2.
 PN
 XX 25-MAY-2001.
 PD
 XX 15-NOV-2000; 2000WO-US31167.
 PF
 XX 15-NOV-1999; 99US-0439735.
 PR
 XX (SCHE) SCHERING CORP.
 PA
 XX Zlot CH, Adema GJ, Figdor C, Phillips JH;
 PI WPI; 2001-367562/38.
 XX N-PSDB; AAD06865.
 DR
 XX New mammalian proteins designated dendritic cell specific transmembrane
 PT protein and DNAX surface protein and the nucleic acids encoding the
 PT polypeptides -
 XX
 XX Claim 2; Page 16; 96pp; English.
 PS
 XX The present sequence is human DNAX surface long form protein (DSP-1L).
 CC The DSP-1L is useful in forensic sciences, e.g. to distinguish
 CC rodent from human or as a marker to distinguish between different
 CC cells exhibiting differential expression or modification patterns.
 CC The DSP-1L and its antibodies are used for preparing kits for use in
 CC molecular biology, immunology or physiology, and in treatment of
 CC conditions associated with abnormal physiology or development, e.g.
 CC allergic conditions (asthma), cancer and autoimmune diseases (diabetes
 CC mellitus). Drug screening using DSP-1L or its fragments can be
 CC performed to identify compounds having binding affinity to or other
 CC relevant isolation effects on the function of DSP-1L. The DSP-1L can
 CC also be used in diagnostic kits and methods for detecting the presence
 CC of another DSP-1L or binding partner.
 XX
 XX Sequence 313 AA;
 SQ
 Query Match 80.0%; Score 32; DB 22; Length 313;
 Best Local Similarity 77.8%; Pred. No. 2.9e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPQLLL 9
 Db 8 LLLPLALL 16
 1:1:1111:11
 1:1:1111:11

RESULT 32
 AAY52588
 ID AAY52588 standard; Protein; 376 AA.
 XX
 XX AAY52588;
 AC
 XX 28-FEB-2000 (first entry)
 DT
 XX Secreted modified Fas ligand, (secFL).
 DE
 XX Fas receptor; Fas ligand; FasL; apoptosis; screening; test compound;
 KW interaction; affinity ligand; identification; purification; therapy;
 KW overstimulation; premature cell death; fulminant liver damage;
 KW understimulation; lymphoproliferative disorders; autoimmune disorder.
 XX
 XX Synthetic.
 OS
 XX Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..211
 FT /note= "Human CD8 secretory signal peptide"
 FT 22..181
 FT Domain

```

FT      /note= "Human CD8 extracellular domain"
FT      184..189
FT      /note= "Hexahistidine motif"
FT      Region
FT      192..197
FT      /note= "Glu-Glu epitope tag"
FT      Domain
FT      198..376
FT      /note= "Human FasL extracellular domain"
XX
XX      US6001962-A.
XX
XX      14-DEC-1999.
XX
XX      15-NOV-1996; 96US-0751512.
XX
XX      15-NOV-1996; 96US-0751512.
XX      (RECC ) UNIV CALIFORNIA.
XX
XX      Ramer JK, Williams LT;
XX
XX      WPI; 2000-061905/05.
XX      N-PSDB; AAZ46253.
XX
XX      Modified Fas ligands for modeling and screening applications,
XX      therapeutic applications and for use as affinity ligands -
XX
XX      Claim 7; Fig 1; 20pp; English.
XX
XX      This sequence represents a secreted modified human Fas ligand (secFL)
XX      which comprises human CD8 secretory sequence and extracellular domain,
XX      heterologous marker sequences (a hexahistidine motif, and a Glu-Glu
XX      epitope tag) and the extracellular domain of Fas ligand (FasL). The
XX      protein is useful for in vitro or in vivo modelling and screening of
XX      test compounds to study the interaction between the Fas receptor and
XX      the Fas ligand and the effects of that interaction on downstream events,
XX      such as Fas mediated apoptosis. The protein may be used as an affinity
XX      ligand for the identification and/or purification of the Fas receptors.
XX      The protein is also useful for therapeutic applications in human and
XX      non-human mammalian patients. Such disorders are characterised by
XX      overstimulation, such as premature cell death and fulminant liver damage,
XX      or understimulation of the Fas mediated pathway, such as
XX      lymphoproliferative disorders and acceleration of autoimmune disorders.
XX
XX      Sequence 376 AA;
XX
XX      Query Match      80.0%; Score 32; DB 21; Length 376;
XX      Best Local Similarity 77.8%; Pred. No. 3.5e+02;
XX      Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX      QY      1 LLLPLQLILL 9
XX            ||||| :||
XX      Db      8 LLLPLALL 16
XX
XX      RESULT 33
XX      AAB92693
XX      ID      AAB92693 standard; Protein; 418 AA.
XX
XX      AC      AAB92693;
XX
XX      DT      26-JUN-2001 (first entry)
XX
XX      Human protein sequence SEQ ID NO:11083.
XX
XX      Human; primer; detection; diagnosis; antisense therapy; gene therapy.
XX
XX      Homo sapiens.
XX
XX      EP1074617-A2.
XX
XX      07-FEB-2001.
XX
XX      28-JUL-2000; 2000EP-0116126.
XX
XX      Query Match      80.0%; Score 32; DB 21; Length 376;
XX      Best Local Similarity 77.8%; Pred. No. 3.5e+02;
XX      Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX      QY      1 LLLPLQLILL 9
XX            ||||| :||
XX      Db      8 LLLPLALL 16
XX
XX      RESULT 34
XX      AAR24863
XX      ID      AAR24863 standard; Protein; 492 AA.
XX
XX      AC      AAR24863;
XX
XX      DT      28-DEC-1992 (first entry)
XX
XX      Sequence of pre-pro stromelysin-3.
XX
XX      Metalloprotease; stromelysin-3; cancer; diagnosis; tumour;
XX      detection.
XX
XX      Mus musculus.
XX
XX      WO9209701-A.
XX
XX      11-JUN-1992.

```

```

XX      29-JUL-1999; 99JP-0248036.
XX      27-AUG-1999; 99JP-0300253.
XX      11-JAN-2000; 2000JP-0118776.
XX      02-MAY-2000; 2000JP-0183767.
XX      09-JUN-2000; 2000JP-0241899.
XX      (HELI-) HELIX RES INST.
XX
XX      Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
XX      Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX      WPI; 2001-318749/34.
XX
XX      primer sets for synthesizing polynucleotides, particularly the 5602
XX      full-length cDNAs defined in the specification, and for the detection
XX      and/or diagnosis of the abnormality of the proteins encoded by the
XX      full-length cDNAs -
XX
XX      Claim 8; SEQ ID 11083; 2537pp + CD ROM; English.
XX
XX      The present invention describes primer sets for synthesising 5602
XX      full-length cDNAs defined in the specification. Where a primer set
XX      comprises: (a) an oligo-dT primer and an oligonucleotide complementary
XX      to the complementary strand of a polynucleotide which comprises one of
XX      the 5602 nucleotide sequences defined in the specification, where the
XX      oligonucleotide comprises at least 15 nucleotides; or (b) a combination
XX      of an oligonucleotide comprising a sequence complementary to the
XX      complementary strand of a polynucleotide which comprises a 5'-end
XX      sequence and an oligonucleotide comprising a sequence complementary to a
XX      polynucleotide which comprises a 3'-end sequence, where the
XX      oligonucleotide comprises at least 15 nucleotides and the combination of
XX      the 5'-end sequence/3'-end sequence is selected from those defined in
XX      the specification. The primer sets can be used in antisense therapy and
XX      in gene therapy. The primers are useful for synthesising polynucleotides,
XX      particularly full-length cDNAs. The primers are also useful for the
XX      detection and/or diagnosis of the abnormality of the proteins encoded by
XX      the full-length cDNAs. The primers allow obtaining of the full-length
XX      cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
XX      AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
XX      AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
XX      represent oligonucleotides, all of which are used in the exemplification
XX      of the present invention.
XX
XX      Sequence 418 AA;
XX
XX      Query Match      80.0%; Score 32; DB 22; Length 418;
XX      Best Local Similarity 77.8%; Pred. No. 3.9e+02;
XX      Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
XX      QY      1 LLLPLQLILL 9
XX            ||||| :||
XX      Db      398 LLLPLPLLL 406
XX
XX      RESULT 34
XX      AAR24863
XX      ID      AAR24863 standard; Protein; 492 AA.
XX
XX      AC      AAR24863;
XX
XX      DT      28-DEC-1992 (first entry)
XX
XX      Sequence of pre-pro stromelysin-3.
XX
XX      Metalloprotease; stromelysin-3; cancer; diagnosis; tumour;
XX      detection.
XX
XX      Mus musculus.
XX
XX      WO9209701-A.
XX
XX      11-JUN-1992.

```


XX 21-NOV-1991; 91WO-FR00924.
 XX 21-NOV-1990; 90GB-0025326.
 XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 PA Basset P, Bellocq JP, Chambon P;
 PI WPI: 1992-217079/26.
 XX N-PSDB; AAQ25699.
 DR Diagnosis and treatment of invasive tumours - especially breast
 PT cancer, using nucleotide probes and antibodies specific for
 PT stromelysin-3
 XX Example; Fig 7; 70pp; French.
 XX A cDNA bank was established from polyA-RNA of a fibroadenoma and
 CC screened with probes derived from polyA-RNA of metastatic tumours.
 CC Clones binding strongly were selected; one of these was found to be
 CC expressed at a high level only in malignant cancers of the breast,
 CC pharynx, head, neck and skin., which are characterised by
 CC deterioration of the extracellular matrix. This clone, which encodes
 CC stromelysin-3 (ST3), was sequenced (AAQ25698). A probe corresponding
 CC to human cDNA encoding ST3 was used to screen a mouse placenta cDNA
 CC library. This led to the isolation of cDNA with the sequence in
 CC AAQ35699. A comparison of the mouse and human ST3 sequences revealed
 CC 89% homology in the amino acids of the mature protein; and
 CC approximately 55% homology in the pre- and pro-domains.
 XX Sequence 492 AA;
 SQ
 Query Match 80.0%; Score 32; DB 13; Length 492;
 Best Local Similarity 77.8%; Pred. NO. 4.6e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPLQLL 9
 Db 15 LLLPLPLL 23
 RESULT 35
 AAQ94928
 ID AAQ94928 standard; Protein; 590 AA.
 AC AAQ94928;
 XX
 DT 16-JUN-2000 (first entry)
 XX Human secreted protein clone pg195_1 protein sequence SEQ ID NO:62.
 DE Human; secreted protein; immunostimulant; immunosuppressant; virucide;
 KW antibacterial; antifungal; cytostatic; antiinflammatory; dermatological;
 KW antidiabetic; antiasthmatic; antiarthritic; antirheumatic; protozoacide;
 KW antithyroid; immune deficiency; severe combined immunodeficiency; SCID;
 KW infection; HIV; hepatitis; malaria; autoimmune disorder; systemic lupus;
 KW connective tissue disease; multiple sclerosis; erythematosis;
 KW rheumatoid arthritis; autoimmune pulmonary inflammation; asthma;
 KW Guillain-Barre syndrome; autoimmune thyroiditis; myasthenia gravis;
 KW insulin dependent diabetes mellitus; graft-versus-host-disease;
 KW autoimmune inflammatory eye disease; allergy.
 XX Homo sapiens.
 OS
 XX WO200009552-A1.
 PN 24-FEB-2000.
 PD 13-AUG-1999; 99WO-US18298.
 PF 14-AUG-1998; 98US-0096622.
 XX 17-AUG-1998; 98US-0096815.

PR 04-SEP-1998; 98US-0099229.
 PR 23-OCT-1998; 98US-0105368.
 PR 08-JAN-1999; 99US-0115234.
 PR 12-FEB-1999; 99US-0119931.
 PR 18-FEB-1999; 99US-0120575.
 PR 30-APR-1999; 99US-0132020.
 PR 11-AUG-1999; 99US-0096622.
 XX (GEMY) GENETICS INST INC.
 XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
 PI Merberg D, Treacy M, Agostino MJ, Steininger RJ, Spaulding V;
 PI Wong GG, Clark HF, Fecthel K;
 DR WPI: 2000-205979/18.
 XX New polynucleotides encoding secreted proteins, which may have e.g.
 PT nutritional, chemokine, immune stimulating or suppressing,
 PT hematopoiesis regulating, tissue growth, activin/inhibin
 PT antiinflammatory or tumor inhibition activity -
 XX Claim 71; Page 534-536; 641pp; English.
 CC AA16618 to AA16697 encode the human secreted proteins given in
 CC AAQ94898 to AAQ94980, isolated from human adult brain, adult thyroid,
 CC adult retina, foetal carcinoma, adult blood, adult neural, foetal kidney,
 CC adult placenta, adult testis, whole embryo, adult cartilage, kidney,
 CC foetal brain, adult thymus, foetal placenta, adult uterus, adult tumour,
 CC and adult bladder, cDNA libraries. The polynucleotides and proteins are
 CC predicted to have biological activities which would make them suitable
 CC for treating, preventing or ameliorating medical conditions in humans
 CC and animals. The polynucleotides can be used as markers for tissues in
 CC which the protein is preferentially expressed, as molecular weight
 CC markers on Southern gels, and as chromosome markers or tags to identify
 CC chromosomes or to map gene positions. The proteins can be used in the
 CC treatment of immune deficiencies and disorders, such as severe combined
 CC immunodeficiency (SCID), as well as viral, bacterial, fungal and other
 CC infections. These infections include human immunodeficiency virus (HIV),
 CC hepatitis, herpesviruses, mycobacteria, Leishmania spp., malaria and
 CC candidiasis. The proteins can be used to treat autoimmune disorders such
 CC as connective tissue disease, multiple sclerosis, systemic lupus
 CC erythematosis, rheumatoid arthritis, autoimmune pulmonary inflammation,
 CC Guillain-Barre syndrome, autoimmunity thyroiditis, insulin dependent
 CC diabetes mellitus, myasthenia gravis, graft-versus-host-disease and
 CC autoimmune inflammatory eye disease. The proteins can also be used to
 CC treat allergic conditions, such as asthma. AA16698 to AA16774 represent
 CC probes for the human secreted proteins from the present invention.
 XX Sequence 590 AA;
 SQ
 Query Match 80.0%; Score 32; DB 21; Length 590;
 Best Local Similarity 77.8%; Pred. NO. 5.5e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLLPLQLL 9
 Db 495 LLLPLQLL 503
 RESULT 36
 AAQ94498
 ID AAQ94498 standard; Protein; 612 AA.
 XX AAQ94498;
 AC
 XX 26-JUN-2001 (first entry)
 DT Human protein sequence SEQ ID NO:15195.
 DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 KW Homo sapiens.
 OS


```

RESULT 38
AAW68155
ID AAW68155 standard; Protein; 1306 AA.
XX AC AAW68155;
XX DT 09-NOV-1998 (first entry)
XX DE Human angiotensin converting enzyme.
XX KW Angiotensin converting enzyme; ACE; hypertension; exercise; human;
XX KW genetic marker.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Peptide 1..29
XX FT /label= sig_peptide
XX FT Protein 30..1307
XX FT /label= Mat_protein
XX FT
XX PN WO9831835-A1.
XX PD 23-JUL-1998.
XX PF 22-DEC-1997; 97WO-US22974.
XX PR 27-MAY-1997; 97US-0048309.
XX PR 16-JAN-1997; 97US-0035382.
XX PA (UYMA-) UNIV MARYLAND BALTIMORE.
XX PA (UYPI-) UNIV PITTSBURGH.
XX PI Ferrell RE, Hagberg JM;
XX DR WPI; 1998-414128/35.
XX DR N-PSDB; AAW41320.
XX PT Analysis of genetic markers to identify subjects who will benefit
XX PT from exercise - also assessing risk of cardiovascular disease from
XX PT angiotensin-converting enzyme genotype
XX PS Disclosure; Page 35-41; 61pp; English.
XX CC This is human angiotensin converting enzyme (ACE). The ACE gene
XX CC (see AAW41320) is polymorphic with 2 common alleles (I and D),
XX CC resulting in 3 genotypes, II, ID and DD. It is an object of the
XX CC invention to identify individuals possessing a certain genotype and
XX CC associated ailment, and to determine if the health of that
XX CC individual can be improved by altering behavior. A claimed method
XX CC comprises identifying individuals having a certain phenotype,
XX CC determining the presence or absence of genetic markers associated
XX CC with the phenotype, and instituting a lifestyle change to exploit
XX CC or counteract the phenotype expressed by the gene marker. If the
XX CC phenotype is hypertension, the gene marker is at least one
XX CC insertion (I) ACE allele and exercise training is instituted to
XX CC decrease systolic and diastolic blood pressure. The gene marker
XX CC can be identified by PCR amplification (see AAW41321-22) of the
XX CC appropriate gene region. The general method can be used to
XX CC identify subjects who will benefit most from physical exercise
XX CC and also to identify those who are likely to be successful in
XX CC sports.
XX SQ Sequence 1306 AA;
Query Match 80.0%; Score 32; DB 19; Length 1306;
Best Local Similarity 77.8%; Pred. No. 1.2e+03;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9
DB 12 LLLPLPLL 20

```

```

RESULT 39
ABB39600
ID ABB39600 standard; Peptide; 34 AA.
XX AC ABB39600;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #7106 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.
XX PT Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human fetal liver -
XX PS Claim 27; SEQ ID NO 32235; 639pp + sequence listing; English.
XX CC The invention relates to a single exon nucleic acid probe for
XX CC measuring human gene expression in a sample derived from human foetal
XX CC liver. The single exon nucleic acid probes may be used for predicting,
XX CC measuring and displaying gene expression in samples derived from human
XX CC foetal liver. The present sequence is a peptide encoded by a single exon
XX CC nucleic acid probe of the invention.
XX CC Note: The sequence data for this patent did not form part of the
XX CC printed specification, but was obtained in electronic format directly
XX CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 34 AA;
Query Match 77.5%; Score 31; DB 22; Length 34;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQLL 9
DB 23 LLLPLVILM 31

```

```

RESULT 40
ABB24303
ID ABB24303 standard; Protein; 34 AA.
XX AC ABB24303;
XX DT 23-JAN-2002 (first entry)
XX DE Protein #6302 encoded by probe for measuring heart cell gene expression.
XX KW Human; gene expression; heart; microarray; vascular system;

```

KW cardiovascular disease; hypertension; cardiac arrhythmia;
 KW congenital heart disease.

XX Homo sapiens.

PN WO200157274-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00666.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488899/53.

XX Single exon nucleic acid probes for analyzing gene expression in human hearts -

XX Claim 15; SEQ ID No 26073; 530pp; English.

XX The present invention relates to single exon nucleic acid probes for measuring human gene expression in a sample derived from human heart (see CCA21535-ABA41305). The present sequence is a protein encoded by one such probe. The probes may be used for predicting, measuring and displaying gene expression in samples derived from the human heart via microarrays. By measuring gene expression, the probes are useful for predicting, diagnosing, grading, staging, monitoring and prognosing diseases of the human heart and vascular system e.g. cardiovascular disease, hypertension, cardiac arrhythmias and congenital heart disease. CC Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO CC at [ftp.wipo.int/pub/published_pct_sequences](http://wipo.int/pub/published_pct_sequences).

XX Sequence 34 AA;

Query Match 77.5%; Score 31; DB 22; Length 34;

Best Local Similarity 77.8%; Pred. No. 49;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

DB 23 LLLPLVILM 31

RESULT 41

AAM60304

ID AAM60304 standard; Protein; 34 AA.

XX AAM60304;

XX 05-NOV-2001 (first entry)

XX Human brain expressed single exon probe encoded protein SEQ ID NO: 32409.

XX Human; brain expressed exon; gene expression analysis; probe;

KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;

KW epilepsy; cancer.

XX Homo sapiens.

XX WO200157275-A2.

XX 09-AUG-2001.

XX

PF 30-JAN-2001; 2001WO-US00667.

PR 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-483446/52.

XX Single exon nucleic acid probes for analyzing gene expression in human brains -

XX Example 4; SEQ ID NO: 32409; 650pp + Sequence Listing; English.

XX The present invention provides a number of single exon nucleic acid probes which are derived from genomic sequences expressed in the human brain. They can be used to measure gene expression in brain cell samples, which may enable the diagnosis and improved treatment of nervous system diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia, epilepsy and cancers. The present sequence is a protein encoded by one of the probes of the invention.

XX Sequence 34 AA;

Query Match 77.5%; Score 31; DB 22; Length 34;

Best Local Similarity 77.8%; Pred. No. 49;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9

DB 23 LLLPLVILM 31

RESULT 42

AAM72938

ID AAM72938 standard; Protein; 34 AA.

XX AAM72938;

XX 06-NOV-2001 (first entry)

XX Human bone marrow expressed probe encoded protein SEQ ID NO: 33244.

XX Human; bone marrow expressed exon; gene expression analysis; probe; microarray; cancer; leukaemia; lymphoma; myeloma.

XX Homo sapiens.

XX WO200157276-A2.

XX 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00668.

XX 04-FEB-2000; 2000US-0180312.

PR 26-MAY-2000; 2000US-0207456.

PR 30-JUN-2000; 2000US-0608408.

PR 03-AUG-2000; 2000US-0632366.

PR 21-SEP-2000; 2000US-0234687.

PR 27-SEP-2000; 2000US-0236359.

PR 04-OCT-2000; 2000GB-0024263.

XX (MOLE-) MOLECULAR DYNAMICS INC.

XX Penn SG, Hanzel DK, Chen W, Rank DR;

XX WPI; 2001-488900/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human bone marrow -
 XX
 XX Example 4; SEQ ID NO: 33244; 658pp + Sequence Listing; English.
 PS
 XX The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC bone marrow. They can be used to measure gene expression in bone marrow
 CC samples, which may enable the improved diagnosis and treatment of cancers
 CC such as lymphoma, leukaemia and myeloma. The present sequence is a
 CC protein encoded by one of the probes of the invention.
 XX
 XX Sequence 34 AA;
 SQ

Query Match 77.5%; Score 31; DB 22; Length 34;
 Best Local Similarity 77.8%; Pred. No. 49;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 DB 23 LLLPLVILM 31
 ||||| ||:
 ||||| ||:

RESULT 43
 AAM19781
 ID AAM19781 standard; Protein; 34 AA.
 XX
 AC AAM19781;
 XX
 XX 12-OCT-2001 (first entry)
 DT
 XX
 DE Peptide #6215 encoded by probe for measuring cervical gene expression.
 XX
 KW Probe; human; microarray; gene expression; cervical epithelial cell;
 KW cervical cancer.
 XX
 OS Homo sapiens.
 XX
 PN WO200157278-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US006070.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488901/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human cervical epithelial cells -
 XX
 PS Claim 27; SEQ ID No 24607; 487pp; English.
 XX
 XX The present invention relates to human single exon nucleic acid probes
 CC (SENPs; see AAI10068-AAI28459). The present sequence is a peptide encoded
 CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
 CC can be used to produce a single exon microarray, which can be used for
 CC measuring human gene expression in a sample derived from human cervical
 CC epithelial cells. By measuring gene expression, the probes are therefore
 CC useful in grading and/or staging of diseases of the cervix, notably

CC cervical cancer.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 XX Sequence 34 AA;
 SQ

Query Match 77.5%; Score 31; DB 22; Length 34;
 Best Local Similarity 77.8%; Pred. No. 49;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 DB 23 LLLPLVILM 31
 ||||| ||:
 ||||| ||:

RESULT 44
 AAM33163
 ID AAM33163 standard; Protein; 34 AA.
 XX
 AC AAM33163;
 XX
 XX 17-OCT-2001 (first entry)
 DT
 XX
 DE Peptide #7200 encoded by probe for measuring placental gene expression.
 XX
 KW Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200157272-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX
 PA (MOLE-) MOLECULAR DYNAMICS INC.
 XX
 PI Penn SG, Hanzel DK, Chen W, Rank DR;
 XX
 DR WPI; 2001-488997/53.
 XX
 PT Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX
 PS Claim 27; SEQ ID No 33432; 654pp; English.
 XX
 XX The present invention relates to single exon nucleic acid probes (SENPs;
 CC see AAI315-AAI37546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 XX Sequence 34 AA;
 SQ

Query Match 77.5%; Score 31; DB 22; Length 34;
 Best Local Similarity 77.8%; Pred. No. 49;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 DB 23 LLLPLVILM 31
 ||||| ||:
 ||||| ||:

RESULT 45
AAG56638
ID AAG56638 standard; Protein; 35 AA.
XX AC AAG56638;
XX DT 18-OCT-2000 (first entry)
DE DE Arabidopsis thaliana protein fragment SEQ ID NO: 72849.
XX DE
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD
XX PF 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135353.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136392.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149315.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.

PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 77.5%; Score 31; DB 21; Length 35;
 Best Local Similarity 77.8%; Pred. No. 50;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLLPLQILL 9
 ||||| :||
 Db 7 LLLPLLLL 15

Search completed: November 6, 2002, 12:05:06
 Job time : 26.7778 secs

THIS PAGE BLANK (USF)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 18.4444 Seconds
(without alignments)
84.413 Million cell updates/sec

Title: US-09-905-083-34
Perfect score: 47
Sequence: 1 VLVNERWVL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriopl.*
- 17: sp_archaea.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	85.1	455	10 Q9FFR7	Q9ffr7 arabidopsis
2	39	83.0	100	6 Q9MYV8	Q9myv8 bos taurus
3	38	80.9	223	5 Q9XY58	Q9xy58 ctienocephal
4	38	80.9	377	10 Q9SNV8	Q9sny8 solanum tub
5	38	80.9	418	10 Q9SNV9	Q9sny9 solanum tub
6	38	80.9	461	5 P91123	P91123 caenorhabdi
7	37	78.7	233	13 Q9PT51	Q9pt51 agkistrodon
8	37	78.7	243	5 Q9VEM7	Q9vem7 drosophila
9	37	78.7	249	13 Q9VEM7	Q9vem7 drosophila
10	37	78.7	249	13 Q9W6K0	Q9w6k0 notothenia
11	37	78.7	250	13 Q93265	Q93265 pseudopleur
12	37	78.7	254	5 Q97098	Q97098 anopheles a
13	37	78.7	255	5 Q97100	Q97100 anopheles d
14	37	78.7	257	13 Q9YGV9	Q9ygv9 agkistrodon
15	37	78.7	257	13 Q9PTL3	Q9ptl3 agkistrodon
16	37	78.7	258	5 Q9XY53	Q9xy53 ctienocephal

17	37	78.7	344	13 Q9W6J9	Q9w6j9 dissostichu
18	37	78.7	675	13 Q9W6J8	Q9w6j8 dissostichu
19	37	78.7	1176	16 Q98IK4	Q98ik4 rhizobium l
20	37	78.7	1421	10 Q49539	Q49539 arabidopsis
21	36	76.6	72	16 Q9PGL9	Q9pgl9 xylella fas
22	36	76.6	111	4 Q9UKR2	Q9ukr2 homo sapien
23	36	76.6	259	5 Q45048	Q45048 anopheles g
24	36	76.6	295	16 Q98N76	Q98n76 rhizobium l
25	36	76.6	345	11 Q923V5	Q923v5 rattus norv
26	36	76.6	350	16 Q98JL3	Q98jl3 rhizobium l
27	36	76.6	426	12 Q89166	Q89166 variola vir
28	36	76.6	426	12 Q57228	Q57228 vaccinia vi
29	36	76.6	426	12 Q9JF80	Q9jfb0 vaccinia vi
30	36	76.6	547	5 Q9VDM6	Q9vdm6 drosophila
31	36	76.6	580	5 Q9VS87	Q9vs87 drosophila
32	36	76.6	3127	5 Q9GT71	Q9gt71 drosophila
33	36	76.6	3497	5 Q9BKA0	Q9bka0 drosophila
34	35	74.5	59	13 Q90WQ4	Q90wq4 paralarabax
35	35	74.5	147	6 Q28523	Q28523 macaca mula
36	35	74.5	178	5 P91101	P91101 caenorhabdi
37	35	74.5	178	13 Q93594	Q93594 dicentrarch
38	35	74.5	237	13 Q91515	Q91515 fugu rubrip
39	35	74.5	238	13 Q9W706	Q9w706 paralichthy
40	35	74.5	240	13 Q98TH0	Q98th0 engraulis j
41	35	74.5	242	13 Q93266	Q93266 pseudopleur
42	35	74.5	242	13 Q9W707	Q9w707 paralichthy
43	35	74.5	242	13 Q92099	Q92099 paranotothe
44	35	74.5	258	6 Q28803	Q28803 pan troglod
45	35	74.5	258	6 Q28804	Q28804 pan troglod

ALIGNMENTS

RESULT 1

Q9FFR7 ID Q9FFR7 PRELIMINARY; PRT; 455 AA.
AC Q9FFR7;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 5, P1 CLONE:MBK23.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=97471969; PubMed=9330910;
RA Sato S., Kotani H., Nakamura Y., Kaneko T., Asamizu E., Fukami M.,
RA Miyajima N., Tabeta S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. I. Sequence
RT features of the 1.6 Mb regions covered by twenty physically assigned
RT P1 clones."
RL DNA Res. 4:215-230(1997).
DR EMBL; AB005233; BABI1469.1; -.
DR InterPro; IPR001810; F-box.
DR Pfam; PF00646; F-box; 1.
SQ SEQUENCE 455 AA; 52305 MW; 763F5D5278D22F13 CRC64;

Query Match 85.1%; Score 40; DB 10; Length 455;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
Db 104 IINERWVL 111

RESULT 2

Q9MYV8

```
ID Q9MYV8 PRELIMINARY; PRT; 100 AA.
AC Q9MYV8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HAPTOGLOBIN (FRAGMENT).
GN HP.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVINDUCT;
RA Lavery K.S., Gabler C., Killian G.J.;
RT "Expression and localization of haptoglobin in the bovine female
RT reproductive tract."; to the EMBL/GenBank/DBJ databases.
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ271156; CAC00531.1; -.
DR HSSP; P00734; 1BA8.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
FT NON_TER 100
FT NON_TER 100
SQ SEQUENCE 100 AA; 11239 MW; PFPEGAESAB43CED9 CRC64;

Query Match 83.08; Score 39; DB 6; Length 100;
Best Local Similarity 75.08; Pred. No. 4.1;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
Db 22 LVNERWLL 29

RESULT 3
Q9XY58 PRELIMINARY; PRT; 223 AA.
AC Q9XY58;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CHYMOTRYPSIN-LIKE SERINE PROTEASE (FRAGMENT).
GN SP-4.
OS Ctenocephalides felis (Cat flea).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Siphonaptera; Pulicidae;
OC Pulicinae; Ctenocephalides.
OX NCBI_TaxID=7515;
RN [1]
RP SEQUENCE FROM N.A.
RA Gaines P.J., Sampson C.M., Rushlow K.E., Stiesler G.L.;
RT "Cloning of a family of serine protease genes from the cat flea
RT Ctenocephalides felis.";
RT Insect Mol. Biol. 81:11-22(1999).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AF053916; AAD21836.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.0PA; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 2.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
FT NON_TER 1
```

```
SQ SEQUENCE 223 AA; 24844 MW; C34EBE9455DD7DE9 CRC64;

Query Match 80.98; Score 38; DB 5; Length 223;
Best Local Similarity 75.08; Pred. No. 15;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 4
Q9SNY8 PRELIMINARY; PRT; 377 AA.
AC Q9SNY8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE BRANCHED-CHAIN AMINO ACID AMINOTRANSFERASE (EC 2.6.1.42).
GN BCAT2.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. RUSSET BURBANK;
RA Campbell M.A., Patel J.K., Meyers J.L., Myrick L.C.;
RT "Isolation of a functional gene encoding for branched-chain amino acid
RT aminotransferase from potato.";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF193846; AAF07192.1; -.
DR HSSP; P00510; 1A3G.
DR InterPro; IPR001544; Aminotran_4.
DR Pfam; PF01063; aminotran_4; 1.
DR ProDom; PD01961; Aminotran_4; 1.
DR PROSITE; PS00770; AA_TRANSFER_CLASS_4; 1.
KW Transferase; Aminotransferase.
SQ SEQUENCE 377 AA; 40924 MW; 8CC8460A5F15B646 CRC64;

Query Match 80.98; Score 38; DB 10; Length 377;
Best Local Similarity 75.08; Pred. No. 26;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWV 8
Db 147 VLVNERWI 154

RESULT 5
Q9SNY9 PRELIMINARY; PRT; 418 AA.
AC Q9SNY9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BRANCHED-CHAIN AMINO ACID AMINOTRANSFERASE (EC 2.6.1.42).
GN BCAT1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. RUSSET BURBANK;
RA Campbell M.A., Patel J.K., Meyers J.L., Myrick L.C.;
RT "Isolation of a functional gene encoding for branched-chain amino acid
RT aminotransferase from potato.";
RT Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF193845; AAF07191.1; -.
DR HSSP; P00510; 1A3G.
```

```
DR InterPro; IPR001544; Aminotran_4.
DR Pfam; PF01063; aminotran_4; 1.
DR ProDom; PD001961; Aminotran_4; 1.
KW Transferase; Aminotransferase.
SQ SEQUENCE 418 AA; 45296 MW; E6A6C2B3A58C814 CRC64;

Query Match 80.9%; Score 38; DB 10; Length 418;
Best Local Similarity 75.0%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 VLVNERWV 8
Db 188 VLVNERWI 195

RESULT 6
P91123
ID P91123 PRELIMINARY; PRT; 461 AA.
AC P91123;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHEITICAL 51.9 KDA PROTEIN.
GN C32E8.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
investigating biology. The C. elegans Sequencing Consortium."
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Gattung S.;
RT "The sequence of C. elegans cosmid C32E8."
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston K.;
RT "Direct Submission."
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88308; AAB42324.2; -.
DR HSSP; P08659; 1LCI.
DR InterPro; IPR000873; AMP-bind.
DR Pfam; PF00501; AMP-binding; 2.
DR PROSITE; PS00455; AMP_BINDING; 1.
KW Hypothetical protein.
SQ SEQUENCE 461 AA; 51902 MW; 8662F7486A0E57A1 CRC64;

Query Match 80.9%; Score 38; DB 5; Length 461;
Best Local Similarity 85.7%; Pred. No. 33;
Matches 6; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 VLVNERW 7
Db 219 VLVNERW 225

RESULT 7
Q9PT51
ID Q9PT51 PRELIMINARY; PRT; 233 AA.
AC Q9PT51;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BEYA-FIBRINOGENASE (FRAGMENT).
```

```
GN BREVINASE.
OS Agkistrodon halys blomhoffi (Mamushi) (Gloydius blomhoffii).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Agkistrodon.
OX NCBI_TaxID=61300;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20302322; PubMed=10845699;
RA Lee J.W., Park W.;
RT "cDNA cloning of brevinase, a heterogeneous two-chain fibrinolytic
enzyme from Agkistrodon blomhoffii brevicaudus snake venom, by serial
hybridization-polymerase chain reaction."
RL Arch. Biochem. Biophys. 377:234-240(2000).
CC 1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL; AJ243757; CAB65936.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.185; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 233 AA; 25725 MW; 1676DC5AF0AB5A42 CRC64;

Query Match 78.7%; Score 37; DB 13; Length 233;
Best Local Similarity 75.0%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
Db 29 LINEWVL 36

RESULT 8
Q9VEM7
ID Q9VEM7 PRELIMINARY; PRT; 243 AA.
AC Q9VEM7;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG4053 PROTEIN.
GN CG4053.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt J.G., Nelson C.R., Miklos G.L.G.,
RA Abrell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
```

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Nushy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weltschbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
RT "The genome sequence of *Drosophila melanogaster*."
RL Science 287:2185-2195(2000).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AE003716; AAF55394.1; -.
DR HSSP; P00766; 1CHG.
DR MEROPS; S01 UPA; -.
DR FlyBase; FBgn0038482; CG4053.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp-SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 243 AA; 26992 MW; 393C4B710563C811 CRC64;

Query Match 78.7%; Score 37; DB 5; Length 243;
Best Local Similarity 55.6%; Pred. No. 26;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9

I:|:|:|
Db 65 VILNEQWIL 73

RESULT 9

Q92046 ID Q92046 PRELIMINARY; PRT; 249 AA.
AC Q92046; Q13261;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PREPROTRYPsin PRECURSOR (EC 3.4.21.4) (ALPHA- AND BETA-TRYPSIN).
OS *Dissostichus mawsoni* (Antarctic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Perciformes;
OC Notothenioidae; Nototheniidae; Dissostichus.
OX NCBI_TaxID=36200;
RN [1]

RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;

RA Chen L., Devries A.L., Cheng C.H.C.;

RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=97268652; PubMed=9108060;

RA Chen L., Devries A.L., Cheng C.H.C.;

RT "Evolution of antifreeze glycoprotein gene from a trypsinogen gene in

RT Antarctic notothenioid fish.";
RL Proc. Natl. Acad. Sci. U.S.A. 94:3811-3816(1997).
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-I-XAA, LYS-I-XAA.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; U58945; AAB57732.1; -.
DR EMBL; U58835; AAB57728.1; -.
DR HSSP; P00763; 1DPO.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp-SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 249 PROTRYPsin.
SQ SEQUENCE 249 AA; 27169 MW; 14F2F0B4F0C6B170 CRC64;

Query Match 78.7%; Score 37; DB 13; Length 249;

Best Local Similarity 66.7%; Pred. No. 26;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9

I:|:|:|
Db 50 VLINQWVL 58

RESULT 10

Q9W6K0 ID Q9W6K0 PRELIMINARY; PRT; 249 AA.
AC Q9W6K0;

DT 01-NOV-1999 (TrEMBLrel. 12, Created)

DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE TRYPSINOGEN-LIKE SERINE PROTEASE.

OS *Notothenia coriiceps* (black rockcod).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Acanthopterygii; Perciformes; Perciformes;

OC Notothenioidae; Nototheniidae; Notothenia.

OX NCBI_TaxID=8208;

RN [1]

RP SEQUENCE FROM N.A.

RA Cheng C.-H.C., Chen L.;

RT "Evolutionary transition from serine protease to antifreeze

glycoprotein captured.";

RL Nature 0:0-0(1999).

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

DR EMBL; AF134323; AAD30107.1; -.
DR HSSP; P00761; 1EPT.

DR MEROPS; S01.151; -.
DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; Tryp-SPC; 1.

DR PROSITE; PS00240; TRYPSIN_DOM; 1.

DR PROSITE; PS00135; TRYPSIN_SER; UNKNOWN_1.

KW Hydrolase; Protease; Serine protease.

SQ SEQUENCE 249 AA; 27419 MW; 69C0D872DFB96ED6 CRC64;

Query Match 78.7%; Score 37; DB 13; Length 249;

Best Local Similarity 66.7%; Pred. No. 26;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9

I:|:|:|
Db 50 VLINQWVL 58

```
RESULT 11
O93265 ID O93265 PRELIMINARY; PRT; 250 AA.
AC O93265;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE TRYPsinNOGEN 1 PRECURSOR (EC 3.4.21.4).
GN TRP1.
OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidei; Pleuronectidae; Pseudopleuronectes.
OX NCBI_TaxID=8265;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RA Douglas S.E., Gallant J.W.;
RT "Isolation of cDNAs for Trypsinogen from the Winter Flounder,
RL J. Mar. Biotechnol. 0:0-0(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
DR EMBL; AF012462; AAC32751.1; -.
DR HSSP; P00761; 1EPT.
DR MEROPS; S01.151; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00135; TRYPsin_SER; UNKNOWN_1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 23 250 TRYPsinNOGEN 1.
SQ SEQUENCE 250 AA; 27466 MW; 99485043B21303FC CRC64;

Query Match 78.7%; Score 37; DB 13; Length 250;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 VLVNERWVL 9
Db 50 VLVNQWVL 58

RESULT 12
O97098 ID O97098 PRELIMINARY; PRT; 254 AA.
AC O97098;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CHYMOTRYPsin 2.
GN CHYM20AA.
OS Anopheles aquasalis.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=42839;
RN [1]
RP SEQUENCE FROM N.A.
RA Almeida R.W., Crisanti A., Mueller H.-M., Ferreira I.I., Galler R.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
DR EMBL; AF051779; AAD17492.1; -.
DR HSSP; P00761; 1EPT.
DR MEROPS; S01.121; -.
DR InterPro; IPR001314; Chymotrypsin.

Query Match 78.7%; Score 37; DB 5; Length 255;
Best Local Similarity 62.5%; Pred. No. 27;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LVNERWVL 9
Db 59 ILNERWIL 66

RESULT 13
O57100 ID O57100 PRELIMINARY; PRT; 255 AA.
AC O57100;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CHYMOTRYPsin 1.
GN CHYM01AD.
OS Anopheles darlingi.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=43151;
RN [1]
RP SEQUENCE FROM N.A.
RA Almeida R.W., Crisanti A., Mueller H.-M., Ferreira I.I., Galler R.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
DR EMBL; AF051781; AAD17494.1; -.
DR HSSP; P00761; 1EPT.
DR MEROPS; S01.121; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 255 AA; 27942 MW; 3E23C8E43061EC90 CRC64;

Query Match 78.7%; Score 37; DB 5; Length 255;
Best Local Similarity 62.5%; Pred. No. 27;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LVNERWVL 9
Db 59 ILNERWIL 66

RESULT 14
O9YGCJ9 ID O9YGCJ9 PRELIMINARY; PRT; 257 AA.
AC O9YGCJ9;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SERINE PROTEASE PRECURSOR.
GN HALY 2.
OS Agkistrodon halys pallás (Chinese water moccasin) (Gloydus halys pallás).
```

```
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 254 AA; 27715 MW; CE0B46D08AB1934C CRC64;

Query Match 78.7%; Score 37; DB 5; Length 254;
Best Local Similarity 62.5%; Pred. No. 27;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LVNERWVL 9
Db 58 ILNERWIL 65

RESULT 13
O57100 ID O57100 PRELIMINARY; PRT; 255 AA.
AC O57100;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CHYMOTRYPsin 1.
GN CHYM01AD.
OS Anopheles darlingi.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=43151;
RN [1]
RP SEQUENCE FROM N.A.
RA Almeida R.W., Crisanti A., Mueller H.-M., Ferreira I.I., Galler R.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
DR EMBL; AF051781; AAD17494.1; -.
DR HSSP; P00761; 1EPT.
DR MEROPS; S01.121; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPsin.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 255 AA; 27942 MW; 3E23C8E43061EC90 CRC64;

Query Match 78.7%; Score 37; DB 5; Length 255;
Best Local Similarity 62.5%; Pred. No. 27;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LVNERWVL 9
Db 59 ILNERWIL 66

RESULT 14
O9YGCJ9 ID O9YGCJ9 PRELIMINARY; PRT; 257 AA.
AC O9YGCJ9;
DT 01-MAY-1999 (TReMBLrel. 10, Created)
DT 01-MAY-1999 (TReMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE SERINE PROTEASE PRECURSOR.
GN HALY 2.
OS Agkistrodon halys pallás (Chinese water moccasin) (Gloydus halys pallás).
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Gloydius.
OX NCBI_TaxID=8714;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=VENOM GLAND;
RA Park D.S., Kim H.D., Chung K.H., Kim D.S., Yun Y.D.;
RT "Molecular Cloning and Characterization of a Novel Plasminogen
activator from Snake Venom";
RL Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL; AF017736; AAD01623.1; -.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.185; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease; Signal.
FT SIGNAL 1 14 POTENTIAL.
FT CHAIN 15 257 SERINE PROTEASE.
SQ SEQUENCE 257 AA; 28180 MW; 7BBFBD7926B3289C CRC64;

Query Match 78.7%; Score 37; DB 13; Length 257;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
Db 53 LINEEWVL 60
RESULT 15
ID Q9PTL3 PRELIMINARY; PRT; 257 AA.
AC Q9PTL3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SALMONASE.
OS Agkistrodon halys brevicaudus (Korean slamosa snake) (Gloydius halys
brevicaudus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Gloydius.
OX NCBI_TaxID=66175;
RN [1]
RP SEQUENCE FROM N.A.
RA Chung K.H., Koh Y.S., Koo B.H., Sohn Y.D., Jang S.Y., Cho S.H.,
RA Kim D.S.;
RT "Salmonase, a two chain direct acting fibrinolytic enzyme from snake
venom";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL; AF176679; AAF25008.1; -.
DR HSSP; P00763; 1DPO.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
SQ SEQUENCE 257 AA; 28921 MW; 4B494A8E6501BE0E CRC64;

Query Match 78.7%; Score 37; DB 13; Length 257;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
Db 53 LINEEWVL 60

RESULT 15
ID Q9PTL3 PRELIMINARY; PRT; 257 AA.
AC Q9PTL3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SALMONASE.
OS Agkistrodon halys brevicaudus (Korean slamosa snake) (Gloydius halys
brevicaudus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Gloydius.
OX NCBI_TaxID=66175;
RN [1]
RP SEQUENCE FROM N.A.
RA Chung K.H., Koh Y.S., Koo B.H., Sohn Y.D., Jang S.Y., Cho S.H.,
RA Kim D.S.;
RT "Salmonase, a two chain direct acting fibrinolytic enzyme from snake
venom";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL; AF176679; AAF25008.1; -.
DR HSSP; P00763; 1DPO.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
SQ SEQUENCE 257 AA; 28921 MW; 4B494A8E6501BE0E CRC64;

Query Match 78.7%; Score 37; DB 13; Length 257;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
Db 53 LINEEWVL 60
RESULT 15
ID Q9PTL3 PRELIMINARY; PRT; 257 AA.
AC Q9PTL3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SALMONASE.
OS Agkistrodon halys brevicaudus (Korean slamosa snake) (Gloydius halys
brevicaudus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Gloydius.
OX NCBI_TaxID=66175;
RN [1]
RP SEQUENCE FROM N.A.
RA Chung K.H., Koh Y.S., Koo B.H., Sohn Y.D., Jang S.Y., Cho S.H.,
RA Kim D.S.;
RT "Salmonase, a two chain direct acting fibrinolytic enzyme from snake
venom";
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL; AF176679; AAF25008.1; -.
DR HSSP; P00763; 1DPO.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease.
SQ SEQUENCE 257 AA; 28921 MW; 4B494A8E6501BE0E CRC64;

Query Match 78.7%; Score 37; DB 13; Length 257;
Best Local Similarity 75.0%; Pred. No. 27;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
Db 53 LINEEWVL 60

RESULT 16
ID Q9XY53 PRELIMINARY; PRT; 258 AA.
AC Q9XY53;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CHYMOTRYPSIN-LIKE SERINE PROTEASE.
GN SP-24.
OS Ctenocephalides felis (Cat flea).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Siphonaptera; Pulicidae;
OC Pulicidae; Ctenocephalides.
OX NCBI_TaxID=7515;
RN [1]
RP SEQUENCE FROM N.A.
RA Gaines P.J., Sampson C.M., Rushlow K.E., Stiegler G.L.;
RT "Cloning of a family of serine protease genes from the cat flea
Insect Mol. Biol. 81:11-22(1999).
RL Insect Mol. Biol. 81:11-22(1999).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL; AF053911; AAD21831.1; -.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.0PA; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Protease; Serine protease.
SQ SEQUENCE 258 AA; 28104 MW; 9063AA6D4904CB48 CRC64;

Query Match 78.7%; Score 37; DB 5; Length 258;
Best Local Similarity 62.5%; Pred. No. 27;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
Db 59 IVNDRWIL 66

RESULT 17
ID Q9W6J9 PRELIMINARY; PRT; 344 AA.
AC Q9W6J9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CHIMERIC ANTIFREEZE GLYCOPROTEIN-SERINE PROTEASE (FRAGMENT).
OS Dissostichus mawsoni (Antarctic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Nototheriidae; Nototheriidae; Dissostichus.
OX NCBI_TaxID=36200;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PANCREAS;
RX MEDLINE=99447034; Pubmed=10519545;

```

RA  Cheng C.H., Chen L.;
RT  "Evolution of an antifreeze glycoprotein.";
RL  Nature 401:443-444(1999).
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC  TRYPsin FAMILY.
DR  EMBL; AF134322; AAD31711.1; -.
DR  HSSP; P00763; 1DPO.
DR  MEROPS; S01.151; -.
DR  InterPro; IPR000104; Antifreeze1.
DR  InterPro; IPR001314; Chymotrypsin.
DR  InterPro; IPR001778; POA_allergen.
DR  InterPro; IPR001254; Trypsin.
DR  Pfam; PF00089; trypsin; 1.
DR  PRINTS; PR00308; ANTIFREEZE1.
DR  PRINTS; PR00722; CHYMOTRYPSIN.
DR  PRINTS; PR00833; POAALLERGEN.
DR  SMART; SM00020; Tryp_SPC; 1.
DR  PROSITE; PS02040; TRYPsin_DOM; 1.
DR  PROSITE; PS00135; TRYPsin_SER; UNKNOWN_1.
KW  Hydrolase; Protease; Serine protease.
FT  NON_TER 1
SQ  SEQUENCE 344 AA; 35353 MW; 8C34DAE49F726CBF CRC64;

Query Match 78.7%; Score 37; DB 13; Length 344;
Best Local Similarity 66.7%; Pred. No. 37;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9
Db 145 VLINQWVL 153

RESULT 18
Q9W6J8
ID Q9W6J8 PRELIMINARY; PRT; 675 AA.
AC Q9W6J8;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CHIMERIC AGF/TRYPsinOGEN-LIKE SERINE PROTEASE (FRAGMENT).
OS Dissostichus mawsoni (Antarctic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidae; Nototheniidae; Dissostichus.
OX NCBI_TaxID=36200;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9947034; PubMed=10519545;
RA Cheng C.H., Chen L.;
RT "Evolution of an antifreeze glycoprotein.";
RL Nature 401:443-444(1999).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
DR  EMBL; AF134321; AAD37247.1; -.
DR  HSSP; P00763; 1DPO.
DR  MEROPS; S01.151; -.
DR  InterPro; IPR000104; Antifreeze1.
DR  InterPro; IPR001314; Chymotrypsin.
DR  InterPro; IPR001778; POA_allergen.
DR  InterPro; IPR001254; Trypsin.
DR  Pfam; PF00089; trypsin; 1.
DR  PRINTS; PR00308; ANTIFREEZE1.
DR  PRINTS; PR00722; CHYMOTRYPSIN.
DR  PRINTS; PR00833; POAALLERGEN.
DR  SMART; SM00020; Tryp_SPC; 1.
DR  PROSITE; PS02040; TRYPsin_DOM; 1.
DR  PROSITE; PS00135; TRYPsin_SER; UNKNOWN_1.
KW Hydrolase; Protease; Serine protease.
FT NON_TER 1
SQ SEQUENCE 675 AA; 63173 MW; A65BD96BA90FE98A CRC64;

Query Match 78.7%; Score 37; DB 13; Length 675;

```

```

Best Local Similarity 66.7%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9
Db 476 VLINQWVL 484

RESULT 19
Q98IK4
ID Q98IK4 PRELIMINARY; PRT; 1176 AA.
AC Q98IK4;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE MLR2360 PROTEIN.
GN MLR2360.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
RW EMBL; AP002999; BAB49512.1; -.
KW Complete proteome.
SQ SEQUENCE 1176 AA; 128152 MW; 882B054E85A4C0AE CRC64;

Query Match 78.7%; Score 37; DB 16; Length 1176;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9
Db 715 MLVNDQWVL 723

RESULT 20
O49539
ID O49539 PRELIMINARY; PRT; 1421 AA.
AC O49539;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PREDICTED PROTEIN.
GN F6H11.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Brandt P., Dose S., Jarke D., Scharfe M., Schon O.,
RA Hoheisel J., Mewes H.W., Mayer K., Schueller C.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL021684; CAAL16682.1; -.
DR HSSP; P00510; 1A3G.
DR InterPro; IPR001544; Aminotran_4.
DR Pfam; PF01063; aminotran_4; 1.

```

DR PRODom; PD001961; AminoTran-4; 1.
 DR PROSITE; PS00770; AA_TRANSFERRIN_CLASS.4; 1.
 SQ SEQUENCE 1421 AA; 163225 MW; C9EL3AD319341588 CRC64;

Query Match 78.7%; Score 37; DB 10; Length 1421;
 Best Local Similarity 66.7%; Pred. No. 1.7e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWYL 9
 |||:||||:
 Db 1208 VLANKRWVI 1216

RESULT 21
 Q9PGL9 PRELIMINARY; PRT; 72 AA.

AC Q9PGL9;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN XF0279.
 GN XF0279.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvarenga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carier H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lenos E.G.M., Lenos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
 RA Machado M.A., Madeira A.M.B.N., Madelira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A.Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silva J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tshako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
 RL Nature 405:151-159(2000).
 DR EMBL; AE003881; AAF8302.1; -;
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 72 AA; 7653 MW; 50019C4BCA1537C2 CRC64;

Query Match 76.6%; Score 36; DB 16; Length 72;
 Best Local Similarity 62.5%; Pred. No. 11;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLVNERWV 8
 :|:||||:
 Db 50 ILMNERWL 57

RESULT 22
 Q9UKR2

ID Q9UKR2 PRELIMINARY; PRT; 111 AA.
 AC Q9UKR2;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE KALLIKREIN-LIKE PROTEIN 5-RELATED PROTEIN 2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yousef G.M., Luo L., Diamandis E.P.;
 RT "Identification of novel human kallikrein-like genes on chromosome
 19q13.3-q13.4.";
 RT Anticancer Res. 79:2843-2852(1999).
 RL [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21121728; PubMed=10675891;
 RA Diamandis E.P., Yousef G.M., Lou L.Y., Magklara A., Obiezu C.;
 RT "The new kallikrein gene family: implications in carcinogenesis.";
 RT Trends Endocrinol. Metab. 11:54-60(2000).
 RL EMBL; AF135025; AAF06086.1; -;
 DR HSSP; P00752; IHIA.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 111 AA; 11984 MW; 38EAB4295635D8D CRC64;

Query Match 76.6%; Score 36; DB 4; Length 111;
 Best Local Similarity 66.7%; Pred. No. 17;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWYL 9
 |||:||||:
 Db 50 VLIDHRWVL 58

RESULT 23
 Q45048 PRELIMINARY; PRT; 259 AA.
 ID Q45048;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE SERINE PROTEINASE.
 OS Anopheles gambiae (African malaria mosquito).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
 OC Anopheles
 OX NCBI_TaxID=7165;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=G3;
 RX MEDLINE=20344664; PubMed=10886405;
 RA Shen Z., Edwards M.J., Jacobs-Lorena M.;
 RT "A gut-specific serine protease from the malaria vector Anopheles
 gambiae is downregulated after blood ingestion.";
 RT Insect Mol. Biol. 9:223-229(2000).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY.
 CC EMBL; AF045250; AAC02700.1; -;
 DR HSSP; P00761; IEPT.
 DR MEROPS; S01.0PA; -;
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.

KW Hydrolase; Serine protease.
SQ SEQUENCE 259 AA; 29271 MW; 094AEE75B4CCE3B6 CRC64;

Query Match. 76.6%; Score 36; DB 5; Length 259;
Best Local Similarity 50.0%; Pred. No. 43;
Matches 4; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
: : : : :
Db 57 IINQRWIL 64

RESULT 24
Q98N76
ID Q98N76 PRELIMINARY; PRT; 295 AA.
AC Q98N76;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE MLO262 PROTEIN.
GN MLO262.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002994; BAB47885.1; -.
DR InterPro; IPR000620; DUF6.
DR Pfam; PF00892; DUF6; 2.
KW Complete proteome.
SQ SEQUENCE 295 AA; 31400 MW; CB8231EBC126CFD3 CRC64;

Query Match 76.6%; Score 36; DB 16; Length 295;
Best Local Similarity 75.0%; Pred. No. 49;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
: : : : :
Db 198 LVSRWVM 205

RESULT 25
Q923Y5
ID Q923Y5 PRELIMINARY; PRT; 345 AA.
AC Q923Y5;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE TRACE AMINE RECEPTOR 4.
GN TA4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY;
RX MEDLINE=21374364; PubMed=11459929;
RA Borowsky B., Adham N., Jones K.A., Raddatz R., Artymyshyn R.,
RA Ogozalek K.L., Durkin M.M., Lakhani P.P., Bonini J.A., Pathirana S.,
RA Boyle N., Pu X., Kouranova E., Lichtblau H., Ochoa F.Y.,
RA Branchek T.A., Gerald C.;

"Trace amines: Identification of a family of mammalian G protein-
RT coupled receptors.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:8966-8971(2001).
DR EMBL; AF380191; AAK71242.1; -.
KW Receptor.
SQ SEQUENCE 345 AA; 38311 MW; DC904127D5B406EC CRC64;

Query Match 76.6%; Score 36; DB 11; Length 345;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
: : : : :
Db 192 VVVNQWVL 200

RESULT 26
Q98JL3
ID Q98JL3 PRELIMINARY; PRT; 350 AA.
AC Q98JL3;
DT 01-OCT-2001 (TREMBlrel. 18, Created)
DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBlrel. 18, Last annotation update)
DE PERMEASE PROTEIN OF RIBOSE ABC TRANSPORTER.
GN MLR1893.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RA "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002998; BAB49152.1; -.
DR InterPro; IPR001851; Bact_bind_dep_transp.
KW Complete proteome.
SQ SEQUENCE 350 AA; 37131 MW; 870722DDBAB22B50 CRC64;

Query Match 76.6%; Score 36; DB 16; Length 350;
Best Local Similarity 71.4%; Pred. No. 59;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWV 8
: : : : :
Db 47 IVNERMI 53

RESULT 27
Q89166
ID Q89166 PRELIMINARY; PRT; 426 AA.
AC Q89166;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE ORF11R.
GN A22R.
OS Variola virus, and
OS variola minor virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10255, 53258;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=Variola virus; STRAIN=GARCIA-1966;
RA Shchelkunov S.N., Totmenin A.V., Sosnovtsev S.V., Safronov P.F.,

RA Resenchuk S.M., Blinov V.M., Sandakhchiev L.S.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RC SPECIES-variola minor virus; STRAIN-GARCIA-1966;
RA Shchelkunov S.N., Totmenin A.V., Gutorov V.V., Safronov P.F.,
RA Massung R.F., Loparev V.N., Knight J.C., Chizhikov V.E., Parsons J.M.,
RA Esposito J.J., Sosnovtsev S.;
RT "Analysis of the complete coding sequence of DNA of alastrim variola
RT minor virus strain Garcia-1966";
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y16780; CAA53893.1; -;
DR EMBL; Y16780; CAB54724.1; -;
SQ SEQUENCE 426 AA; 49173 MW; 4289789FAC5DF621 CRC64;

Query Match 76.6%; Score 36; DB 12; Length 426;
Best Local Similarity 55.6%; Pred. No. 72;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9
||:| |::
Db 130 VLLNNRWIM 138

RESULT 28

ID 057228 PRELIMINARY; PRT; 426 AA.

AC 057228;

DT 01-JUN-1998 (TREMBlrel. 06, Created)

DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE PUTATIVE 49.1K PROTEIN.

GN MVAL32R

OS Vaccinia virus (strain Ankara).

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

OC Orthopoxvirus.

OX NCBI_TaxID=126794;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ANKARA;

RA Antoine G., Scheiflinger F., Falkner F.G., Dörner F.;

RT "The complete genomic sequence of the Modified Vaccinia Ankara (MVA)

RT strain.";

RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; U94848; AAB96523.1; -;

SQ SEQUENCE 426 AA; 49074 MW; 650D8EE38CC76A78 CRC64;

Query Match 76.6%; Score 36; DB 12; Length 426;
Best Local Similarity 55.6%; Pred. No. 72;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9
||:| |::
Db 130 VLLNNRWIM 138

RESULT 29

ID 09JF80 PRELIMINARY; PRT; 426 AA.

AC 09JF80;

DT 01-OCT-2000 (TREMBlrel. 15, Created)

DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)

DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)

DE TAZ1R.

OS Vaccinia virus (strain Tian Tan).

OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;

OC Orthopoxvirus.

OX NCBI_TaxID=10253;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-TIAN TAN;

RA Jin Q., Hou Y.D., Cheng N.H., Yao E.M., Cheng S.X., Yang X.K.,

RA Jing D.Y., Yu W.H., Yuan J.S., Ma X.J.;

RT "Complete genomic sequence of vaccinia virus (Tian Tan strain).";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF095689; AAF34015.1; -;

SQ SEQUENCE 426 AA; 49158 MW; 66853F628A9A896A CRC64;

Query Match 76.6%; Score 36; DB 12; Length 426;
Best Local Similarity 55.6%; Pred. No. 72;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9
||:| |::
Db 130 VLLNNRWIM 138

RESULT 30

Q9VDW6

ID Q9VDW6 PRELIMINARY; PRT; 547 AA.

AC Q9VDW6;

DT 01-MAY-2000 (TREMBlrel. 13, Created)

DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)

DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)

DE DYSTROPHIN PROTEIN.

GN DYS OR DYSTROPHIN OR CG7240 OR CG7243.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt J., Nelson C.R., Miklos G.L.G.,

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reiner K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster";

RL Science 287:2185-2195(2000).

DR EMBL; AE003726; AAF55673.1; -;

DR HSSP; P46939; 1BHD.

DR FlyBase; FBgn0024242; Dys.

DR InterPro: IPR001589; Actinin_act_bind.
 DR InterPro: IPR001715; Calponin_hom.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF00307; CH; 2.
 DR SMART: SM00033; CH; 2.
 DR SMART: SM00150; SPEC; 2.
 DR PROSITE: PS00019; ACTININ_1; 1.
 DR PROSITE: PS50021; CH; 2.
 SQ SEQUENCE 547 AA; 61893 MW; 183E801D5BE1A8F CRC64;

Query Match 76.6%; Score 36; DB 5; Length 547;

Best Local Similarity 85.7%; Pred. No. 95;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLVNERW 7

Db 405 VLLNERW 411

RESULT 31

Q9VS87 PRELIMINARY; PRT; 580 AA.

AC Q9VS87;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CG8528 PROTEIN.
 GN CG8528.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt E.G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.F., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
 RA Fowler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Godek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laspo K., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL: AE003558; AAF50540.1; -.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.UPA; -.
 DR FlyBase: FBgn0035796; CG8528.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin; 2.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; Tryp_Spc; 2.
 DR PROSITE: PS02040; TRYPsin_DOM; 2.
 DR PROSITE: PS00135; TRYPsin_SER; 2.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 580 AA; 65341 MW; 7AB43069B80F7E12 CRC64;

Query Match 76.6%; Score 36; DB 5; Length 580;

Best Local Similarity 55.6%; Pred. No. 1e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9

Db 383 VILNRRWIL 391

RESULT 32

Q9GT71 PRELIMINARY; PRT; 3127 AA.

AC Q9GT71;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE DYSTROPHIN-LIKE PROTEIN DYS.
 GN DYS OR CG7240 OR CG7243.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RX PubMed=11018515;

RA Greener M.J., Roberts R.G.;

RT "Conservation of components of the dystrophin complex in

RL Drosophila(1).";

RL FEBS Lett. 482:13-18(2000).

DR EMBL: AF277386; AAG17395.1; -.
 DR HSSP: P46939; 1BHD.

DR FlyBase: FBgn002424; Dys.

DR InterPro: IPR001589; Actinin_act_bind.

DR InterPro: IPR001715; Calponin_hom.

DR InterPro: IPR003716; RNA_pol_omega.

DR InterPro: IPR002017; Spectrin.

DR InterPro: IPR001202; WW.

DR InterPro: IPR000433; ZnF_ZZ.

DR Pfam: PF00307; CH; 2.

DR Pfam: PF00435; spectrin; 13.

DR Pfam: PF00569; ZZ; 1.

DR SMART: SM00033; CH; 2.

DR SMART: SM00150; SPEC; 11.

DR SMART: SM00456; WW; 1.

DR SMART: SM00291; ZnF_ZZ; 1.

DR PROSITE: PS00019; ACTININ_1; 1.

DR PROSITE: PS50021; CH; 2.

DR PROSITE: PS50020; WW_DOMAIN_2; 1.

SQ SEQUENCE 3127 AA; 358057 MW; D217E7891EDE6C96 CRC64;

Query Match

Best Local Similarity 76.6%; Score 36; DB 5; Length 3127;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

GN C24G7.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson R., Alnsough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA Shalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RL Nature 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Goela D., Wilson R.;
RT "the sequence of C. elegans cosmid C24G7.";
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; U88310; AAB42339.1; -;
SQ SEQUENCE 178 AA; 20565 MW; DE8DF12BB569E875 CRC64;

Query Match 74.5%; Score 35; DB 5; Length 178;
Best Local Similarity 55.6%; Pred. No. 44;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 VLVNERWVYL 9
:::|||||
Db 20 IIEENEKVL 28

RESULT 37
O93594 ID O93594 PRELIMINARY; PRT; 178 AA.
AC O93594;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRYPsin (EC 3.4.21.4) (FRAGMENT).
OS Dicertrarchus labrax (European sea bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Moronidae; Dicertrarchus.
OX NCBI_TaxID=13489;
RN [1]
RP SEQUENCE FROM N.A.
RA Peres A., Zambonino Infante J.L., Cahu C.L.;
RT "Dietary regulation of activities and mRNA levels of trypsin and
RT amylase in sea bass (Dicertrarchus labrax) larvae.";
RL Fish Physiol. Biochem. 19:145-152(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
DR EMBL; AJ006882; CAA07315.1; -;
DR HSSP; P35031; IBIT.
DR MEROPS; S01.151; -;
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
Qy Pfam; PF00089; trypsin; 1.
Db Pfam; PF00089; trypsin; 1.

Query Match 74.5%; Score 35; DB 13; Length 237;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVYL 9
|||||:::
Db 44 LVNENWV 51

RESULT 39
O9W706 ID O9W706 PRELIMINARY; PRT; 238 AA.
AC O9W706;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRYPsinOGEN 2 (FRAGMENT).
OS Paralichthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
FT NON_TER 178
SQ SEQUENCE 178 AA; 19352 MW; 13F13EBEC80EDC57 CRC64;

Query Match 74.5%; Score 35; DB 13; Length 178;
Best Local Similarity 75.0%; Pred. No. 44;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVYL 9
|||||:::
Db 15 LVNENWV 22

RESULT 38
O91515 ID O91515 PRELIMINARY; PRT; 237 AA.
AC O91515;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRYPsinOGEN (FRAGMENT).
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang K., Gan L., Lee I., Roach J., Hood L.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY.
DR EMBL; U25747; AAA75001.1; -;
DR HSSP; P35031; IBIT.
DR MEROPS; S01.151; -;
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
FT NON_TER 178
SQ SEQUENCE 237 AA; 25726 MW; 30D2DBAAC39080C2 CRC64;

Query Match 74.5%; Score 35; DB 13; Length 237;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVYL 9
|||||:::
Db 44 LVNENWV 51

RESULT 39
O9W706 ID O9W706 PRELIMINARY; PRT; 238 AA.
AC O9W706;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRYPsinOGEN 2 (FRAGMENT).
OS Paralichthys olivaceus (Flounder).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 CC Pleuronectoidei; Paralichthyidae; Paralichthys.
 OX NCBI_TaxID=8255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=PANCREAS;
 RA Suzuki T., Srivastava A.S., Kurokawa T.;
 RT "Japanese flounder mRNA for trypsinogen 2.";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; AB029751; BAAB2363.1; -.
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.151; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS02040; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease.
 FT NON_TER 1
 SQ SEQUENCE 238 AA; 26071 MW; F2B8908085B8D062 CRC64;

 Query Match 74.5%; Score 35; DB 13; Length 238;
 Best Local Similarity 75.0%; Pred. No. 61;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 Qy 2 LVNERWVL 9
 Db 46 LVNENWV 53
 |||||
 |||||

 RESULT 40
 Q98TH0 PRELIMINARY; PRT; 240 AA.
 AC Q98TH0;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE TRYPSINOGEN.
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Clupeomorpha; Engraulidae;
 CC Engraulis.
 OX NCBI_TaxID=42892;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Watabe S., Ahsan M.N., Funabara D.;
 RT "Anchovy trypsinogen mRNA.";
 RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; AB041929; BAB40329.1; -.
 DR HSSP; P00763; IDPO.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS02040; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 240 AA; 26026 MW; 299B119BFF071464 CRC64;

 Query Match 74.5%; Score 35; DB 13; Length 240;
 Best Local Similarity 75.0%; Pred. No. 61;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
 Db 48 LVNENWV 55
 |||||
 |||||

 RESULT 41
 Q93Z66 PRELIMINARY; PRT; 242 AA.
 AC Q93Z66;
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE TRYPSINOGEN 2 PRECURSOR (EC 3.4.21.4).
 GN TRP2.
 OS Pseudopleuronectes americanus (Winter flounder) (Pleuronectes americanus).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 CC Pleuronectoidei; Pleuronectidae; Pseudopleuronectes.
 OX NCBI_TaxID=8265;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=INTESTINE, PYLORIC CAECA, AND PANCREAS;
 RA Douglas S.E., Gallant J.W.;
 RT "Isolation of cDNAs for Trypsinogen from the Winter Flounder, Pleuronectes americanus.";
 RL J. Mar. Biotechnol. 0:0-0(1998).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; AF012463; AAC32752.1; -.
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.151; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_Spc; 1.
 DR PROSITE; PS02040; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Signal.
 FT SIGNAL 1 15 POTENTIAL.
 FT CHAIN 21 242 TRYPSINOGEN 2
 SQ SEQUENCE 242 AA; 26180 MW; 08D2A834FB289080 CRC64;

 Query Match 74.5%; Score 35; DB 13; Length 242;
 Best Local Similarity 75.0%; Pred. No. 62;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
 Db 49 LVNENWV 56
 |||||
 |||||

 RESULT 42
 Q9W7Q7 PRELIMINARY; PRT; 242 AA.
 AC Q9W7Q7;
 DT 01-NOV-1999 (TREMBLrel. 12, Created)
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE TRYPSINOGEN 1.
 OS Paralichthys olivaceus (Flounder).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 CC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 CC Pleuronectoidei; Paralichthyidae; Paralichthys.
 OX NCBI_TaxID=8255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA TISSUE=PANCREAS;

RA Suzuki T., Srivastava A.S., Kurokawa T.,
RT "Japanese flounder mRNA for trypsinogen 1,"
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC -!- TRYPSIN FAMILY.
DR EMBL; AB029750; BAA82362.1; -.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.I51; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 242 AA; 26548 MW; 6DA722C80BC194A2 CRC64;

Query Match 74.5%; Score 35; DB 13; Length 242;
Best Local Similarity 75.0%; Pred. No. 62;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
|||||
Db 49 LVNENWV 56

RESULT 43
Q92099 PRELIMINARY; PRT; 242 AA.
AC Q92099;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRYPSIN PRECURSOR (EC 3.4.21.4).
OS Paranotothenia magellanica (Maori cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes;
OC Notothenioidei; Nototheniidae; Paranotothenia.
OX NCBI_TaxID=37005;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=PYLORIC CARCA;
RX MEDLINE=97104330; PubMed=8948488;
RA Genicot S., Rentier-Deirue F., Edwards D., Van Beeumen J., Gerday C.;
RT "Trypsin and trypsinogen from an antarctic fish: molecular basis of
cold adaptation.";
RL Biochim. Biophys. Acta 1298:45-57(1996).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC -!- TRYPSIN FAMILY.
DR EMBL; X82223; CAA57701.1; -.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.I51; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 13
SQ SEQUENCE 242 AA; 26201 MW; 3F4DE7CE90C4477C CRC64;

Query Match 74.5%; Score 35; DB 13; Length 242;
Best Local Similarity 75.0%; Pred. No. 62;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
|||||
Db 49 LVNENWV 56

RESULT 44
Q28803 PRELIMINARY; PRT; 258 AA.
AC Q28803;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HAPTOGLOBIN (FRAGMENT).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89008487; PubMed=3170608;
RA McEvoy S., Maeda N.;
RT "Complex events in the evolution of the haptoglobin gene cluster in
primates.";
RL J. Biol. Chem. 263:15740-15747(1988).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC -!- TRYPSIN FAMILY.
DR EMBL; M20760; AAA35412.1; -.
DR HSSP; P00763; LDPO.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 258 AA; 28643 MW; 05063CF3F34A2E33 CRC64;

Query Match 74.5%; Score 35; DB 6; Length 258;
Best Local Similarity 62.5%; Pred. No. 66;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
|||||
Db 43 LINEQWLL 50

RESULT 45
Q28804 PRELIMINARY; PRT; 258 AA.
AC Q28804;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HAPTOGLOBIN (FRAGMENT).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89008487; PubMed=3170608;
RA McEvoy S., Maeda N.;
RT "Complex events in the evolution of the haptoglobin gene cluster in
primates.";
RL J. Biol. Chem. 263:15740-15747(1988).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC -!- TRYPSIN FAMILY.
DR EMBL; M20761; AAA35413.1; -.
DR HSSP; P08709; IFAK.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.

Db 49 LVNENWV 56

RESULT 44
Q28803 PRELIMINARY; PRT; 258 AA.
AC Q28803;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HAPTOGLOBIN (FRAGMENT).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89008487; PubMed=3170608;
RA McEvoy S., Maeda N.;
RT "Complex events in the evolution of the haptoglobin gene cluster in
primates.";
RL J. Biol. Chem. 263:15740-15747(1988).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC -!- TRYPSIN FAMILY.
DR EMBL; M20760; AAA35412.1; -.
DR HSSP; P00763; LDPO.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 258 AA; 28643 MW; 05063CF3F34A2E33 CRC64;

Query Match 74.5%; Score 35; DB 6; Length 258;
Best Local Similarity 62.5%; Pred. No. 66;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
|||||
Db 43 LINEQWLL 50

RESULT 45
Q28804 PRELIMINARY; PRT; 258 AA.
AC Q28804;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HAPTOGLOBIN (FRAGMENT).
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89008487; PubMed=3170608;
RA McEvoy S., Maeda N.;
RT "Complex events in the evolution of the haptoglobin gene cluster in
primates.";
RL J. Biol. Chem. 263:15740-15747(1988).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC -!- TRYPSIN FAMILY.
DR EMBL; M20761; AAA35413.1; -.
DR HSSP; P08709; IFAK.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.

DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 KW Hydrolase; Serine protease.
 FT NON_TER 1
 SQ SEQUENCE 258 AA; 28589 MW; 858A7106DEBC778D CRC64;

Query Match 74.5%; Score 35; DB 6; Length 258;
 Best Local Similarity 62.5%; Pred. No. 66;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Oy 2 LVNERWVL 9
 I::I::I
 Db 43 LINEQWLL 50

Search completed: November 6, 2002, 12:12:04
 Job time : 21.4444 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 5.33333 Seconds
(without alignments)
65.339 Million cell updates/sec

Title: US-09-905-083-34
Perfect score: 47
Sequence: 1 VLVNERWVL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	100.0	253	1 KLK7_HUMAN	P49862 homo sapien
2	38	80.9	232	1 KLK_PIG	P00752 sus scrofa
3	38	80.9	260	1 NRPN_MOUSE	O61955 mus musculus
4	38	80.9	260	1 NRPN_RAT	O88780 rattus norv
5	37	78.7	250	1 TRYP_PLEPL	P35034 pleuronecte
6	37	78.7	257	1 VSP4_TRIMU	Q91510 trimeresuru
7	37	78.7	257	1 VSP5_TRIMU	Q91511 trimeresuru
8	37	78.7	257	1 VSP7_TRIMU	Q94984 trimeresuru
9	37	78.7	257	1 VSPC_TRIGA	O13062 trimeresuru
10	37	78.7	258	1 CTR2_ANOGA	Q17025 anopheles g
11	37	78.7	259	1 CTR1_ANOGA	Q27289 anopheles g
12	37	78.7	260	1 VSP1_TRIFE	Q94968 trimeresuru
13	37	78.7	260	1 VSP2_TRIFL	O13057 trimeresuru
14	36	76.6	248	1 KLKC_HUMAN	Q94980 homo sapien
15	36	76.6	258	1 GRAM_RAT	Q03238 rattus norv
16	36	76.6	260	1 ESTA_CANFA	P09582 canis famil
17	36	76.6	426	1 VA20_VACCC	P20995 vaccinia vi
18	36	76.6	426	1 VA20_VARV	P33843 variola vir
19	36	76.6	516	1 Y4BL_RHISN	P55379 rhizobium s
20	35	74.5	231	1 TRY2_SALSA	P35032 salmo salar
21	35	74.5	242	1 TRY1_SALSA	P35031 salmo salar
22	35	74.5	247	1 TRY1_HUMAN	P07477 homo sapien
23	35	74.5	253	1 CFAD_HUMAN	P00746 homo sapien
24	35	74.5	257	1 VSP1_TRIMU	Q91507 trimeresuru
25	35	74.5	257	1 VSP2_TRIMU	Q91508 trimeresuru
26	35	74.5	257	1 VSP3_AGKAC	Q918x0 agkistrodon
27	35	74.5	257	1 VSP3_TRIMU	Q91509 trimeresuru
28	35	74.5	274	1 FA9_SHEEP	P16291 ovis aries
29	35	74.5	274	1 RS2_AQUAE	O67809 aquifex aco
30	35	74.5	324	1 TEST_MOUSE	Q91507 mus musculus
31	35	74.5	329	1 HPT_CANFA	P19006 canis famil
32	35	74.5	347	1 HPT1_HUMAN	P00737 homo sapien
33	35	74.5	347	1 HPT_ATEGE	P50417 ateles geof

RESULT 1				
ID	KLK7_HUMAN	STANDARD;	PRT;	253 AA.
AC	P49862;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic enzyme) (hSCCE).			
DE	KLK7 OR PRSS6 OR SCCE.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RA	Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A., Egelrud T.;			
RT	"Cloning, expression, and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase.";			
RL	J. Biol. Chem. 269:19420-19426(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Keratinocytes;			
RA	Yousef G.M., Scorillas A., Diamandis E.P.;			
RT	"Molecular characterization, mapping and tissue expression of the human stratum corneum chymotryptic enzyme gene.";			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20510030; PubMed=11054574;			
RA	Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J., Moss P., Paepker B., Wang K.;			
RT	"Sequencing and expression analysis of the serine protease gene cluster located in chromosome 19q13 region.";			
RL	Gene 257:119-130(2000).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=95314630; PubMed=7794273;			
RA	Skytt A., Stroemqvist M., Egelrud T.;			
RT	"Primary substrate specificity of recombinant human stratum corneum chymotryptic enzyme.";			
RL	Biochem. Biophys. Res. Commun. 211:586-589(1995).			
CC	-!- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE CLEAVES INSULIN B CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE-1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION OF PRECURSORS TO INFLAMMATORY CYTOKINES.			
CC	-!- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE ALSO SEEN IN THE BRAIN AND KIDNEY.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			

34	35	74.5	348	1	HPTR_HUMAN	P00739 homo sapien
35	35	74.5	406	1	HP2_HUMAN	P00738 homo sapien
36	35	74.5	416	1	FA9_BOVIN	P00741 bos taurus
37	35	74.5	421	1	ACRO_HUMAN	P10323 homo sapien
38	35	74.5	688	1	C1S_HUMAN	P09871 homo sapien
39	34	72.3	178	1	YOAC_BACSU	P45900 bacillus su
40	34	72.3	228	1	VSPA_LACMU	P33589 lachesis mu
41	34	72.3	231	1	VSP1_AGKCO	P09872 agkistrodon
42	34	72.3	232	1	VSP1_BOTJA	P81824 bothrops ja
43	34	72.3	232	1	VSPA_BOTJA	P81861 bothrops ja
44	34	72.3	235	1	VSP2_AGKBI	Q95933 agkistrodon
45	34	72.3	238	1	VSP1_AGKHA	P81176 agkistrodon

ALIGNMENTS

```

CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L33404; AAC37551.1; -.
DR EMBL; AF166330; AAD49718.1; -.
DR EMBL; AF243527; AAG33360.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.300; -.
DR MIM; 604438; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal.
FT SIGNAL 1 22
FT PROPEP 23 29
FT CHAIN 30 253
FT ACT_SITE 112 112
FT ACT_SITE 205 205
FT DISULFID 36 137
FT DISULFID 55 71
FT DISULFID 137 239
FT DISULFID 144 211
FT DISULFID 176 190
FT DISULFID 201 226
FT CARBOHYD 246 246
FT SEQUENCE 253 AA; 27525 MW; 2D68B6B15A76A668 CRC64;
Query Match 100.08; Score 47; DB 1; Length 253;
Best Local Similarity 100.08; Pred. No. 0.092;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLVNERWVL 9
D 58 VLVNERWVL 66
|||||||
RESULT 2
KLK_PIG STANDARD; PRT; 232 AA.
AC P00752;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glandular kallikrein (EC 3.4.21.35) (Tissue kallikrein).
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE.
RC TISSUE=Pancreas;
RA Tschesche H., Mair G., Codec G., Fiedler F., Ehret W., Hirschauer C.,
RA Lemon M., Fritz H., Schmidt-Kastner G., Kutzbach C.;
RA "The primary structure of porcine glandular kallikreins.";
RA Adv. Exp. Med. Biol. 120:245-260(1979).
RN [2]
RP X-RAY CRYSTALLOGRAPHY (2 ANGSTROMS), AND REVISIONS.
RX MEDLINE=83189107; PubMed=6551452;
RA Bode W., Chen Z., Bartels K., Kutzbach C., Schmidt-Kastner G.,
RA Bartunik H.;

```

```

RT "Refined 2-A X-ray crystal structure of porcine pancreatic kallikrein
RT A, a specific trypsin-like serine proteinase. Crystallization,
RT structure determination, crystallographic refinement, structure and
RT its comparison with bovine trypsin.";
RL J. Mol. Biol. 164:237-282(1983).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF COMPLEX WITH HIRUSTATIN.
RX MEDLINE=97184690; PubMed=9032072;
RA Mittl P.R., di Marco S., Fendrich G., Pohlig G., Heim J.,
RA Sommerhoff C., Fritz H., Priestle J.P., Grutter M.G.;
RA "A new structural class of serine protease inhibitors revealed by the
RA structure of the hirustasin-kallikrein complex.";
RL Structure 5:253-264(1997).
RN [4]
RP ERRATUM.
RA Mittl P.R., di Marco S., Fendrich G., Pohlig G., Heim J.,
RA Sommerhoff C., Fritz H., Priestle J.P., Grutter M.G.;
RL Structure 5:585-585(1997).
RN [5]
RP STRUCTURE OF CARBOHYDRATES.
RX MEDLINE=89062455; PubMed=3196708;
RA Tomiya N., Yameguchi T., Awaya J., Kurono M., Endo S., Arata Y.,
RA Takahashi N., Ishihara H., Mori M., Tejima S.;
RA "Structural analyses of asparagine-linked oligosaccharides of porcine
RA pancreatic kallikrein.";
RL Biochemistry 27:7146-7154(1988).
CC -!- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of Arg-|-Xaa bonds in
CC small molecule substrates. Highly selective action to release
CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
CC Met-|-Xaa or Leu-|-Xaa.
CC -!- SUBUNIT: HETERODIMER OF TWO CHAINS, A AND B, HELD BY A DISULFIDE
CC BOND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.
PTR; A00938; KQPG.
PDB; 2PKA; 19-JUL-84.
PDB; 2KAI; 15-OCT-91.
PDB; 1HIA; 24-DEC-97.
DR MEROPS; S01.160; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS0240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Pancreas; 3D-structure.
FT CHAIN 1 80
FT CHAIN 81 232
FT SITE 79 82
FT DISULFID 7 144
FT DISULFID 26 42
FT DISULFID 121 190
FT DISULFID 155 169
FT DISULFID 180 205
FT CARBOHYD 78 78
FT CARBOHYD 225 225
FT ACT_SITE 41 41
FT ACT_SITE 89 89
FT ACT_SITE 184 184
FT STRAND 2 2
FT TURN 3 3
FT STRAND 5 6
FT TURN 9 10
FT TURN 13 14
FT STRAND 15 20
FT TURN 21 22
FT STRAND 23 32
FT TURN 33 34
FT STRAND 35 38
N-LINKED (GLCNAC...).
N-LINKED (GLCNAC...).
CHARGE RELAY SYSTEM.
CHARGE RELAY SYSTEM.
CHARGE RELAY SYSTEM.

```

FT HELIX 40 42
 FT STRAND 48 51
 FT STRAND 55 55
 FT TURN 56 57
 FT STRAND 64 66
 FT STRAND 68 73
 FT TURN 75 76
 FT TURN 83 84
 FT STRAND 91 95
 FT TURN 117 118
 FT STRAND 120 125
 FT STRAND 141 141
 FT STRAND 143 150
 FT HELIX 152 155
 FT TURN 156 158
 FT TURN 165 166
 FT STRAND 167 171
 FT TURN 173 174
 FT STRAND 178 178
 FT TURN 181 182
 FT TURN 184 185
 FT STRAND 187 190
 FT TURN 191 192
 FT STRAND 193 200
 FT TURN 205 206
 FT TURN 208 209
 FT STRAND 212 216
 FT HELIX 217 230
 SQ SEQUENCE 232 AA; 25589 MW; 012DEE322060C334 CRC64;

Query Match 80.9%; Score 38; DB 1; Length 232;

Best Local Similarity 77.8%; Pred. No. 4;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

||||:|

DB 29 VLVNPKRWVL 37

RESULT 3

ID NRPN_MOUSE STANDARD; PRT; 260 AA.
 AC O61955;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Neuropsin precursor (EC 3.4.21.-) (NP) (Kallikrein 8).
 GN KLK8 OR PRSS19 OR NRPN.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Hippocampus;
 RX MEDLINE=95348817; PubMed=7623137;
 RA Chen Z.-L., Yoshida S., Kato K., Momota Y., Suzuki J., Tanaka T.,
 RA Ito J., Nishino H., Aimoto S., Kiyama H., Shiosaka S.;
 RT "Expression and activity-dependent changes of a novel limbic-serine
 RT protease gene in the hippocampus.";
 RL J. Neurosci. 15:5088-5097(1995).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Yoshida S., Hirata A., Inoue N., Shiosaka S.;
 RT "Cloning and assignment of mouse neuropsin gene, Prss19 to chromosome
 RT 7B4.";
 RN [3]
 RP Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF N-TERMINUS, AND CHARACTERIZATION.
 RC STRAIN=BALB/C; TISSUE=Brain;
 RX MEDLINE=98225202; PubMed=9556608;
 RA Shimizu C., Yoshida S., Shibata M., Kato K., Momota Y., Matsumoto K.,
 RA Shiosaka T., Midorikawa R., Kamachi T., Kawabe A., Shiosaka S.;

RT "Characterization of recombinant and brain neuropsin, a
 RT plasticity-related serine protease.";
 RL J. Biol. Chem. 273:11189-11196(1998).
 RN [4]
 RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 33-257.
 RC TISSUE=Hippocampus;
 RX MEDLINE=99134351; PubMed=9933620;
 RA Kishi T., Kato M., Shimizu T., Kato K., Matsumoto K., Yoshida S.,
 RA Shiosaka S., Hakoshima T.;
 RT "Crystal structure of neuropsin, a hippocampal protease involved in
 RT kindling epileptogenesis.";
 RL J. Biol. Chem. 274:4220-4224(1999).
 CC -!- FUNCTION: SUGGESTED TO BE INVOLVED IN KINDLING EPILEPTOGENESIS AND
 CC HIPPOCAMPAL PLASTICITY. HAS A STRONG PROTEOLYTIC ACTIVITY AGAINST
 CC FIBRONECTIN.
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
 CC -!- ENZYME REGULATION: STRONGLY INHIBITED BY DIISOPROPYL
 CC FLUOROPHOSPHATE, LEUPEPTIN AND (4-AMIDINOPHENYL)METHANESULFONYL 1-
 CC FLUORIDE.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE LIMBIC SYSTEM OF
 CC MOUSE BRAIN AND IS LOCALIZED AT HIGHEST CONCENTRATION IN PYRAMIDAL
 CC NEURONS OF THE HIPPOCAMPAL CA1-3 SUBFIELDS.
 CC -!- MASS SPECTROMETRY: MW=26613; METHOD=MALDI; RANGE=29-260.
 CC -!- MASS SPECTROMETRY: MW=26229; METHOD=MALDI; RANGE=33-260.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: D30785; BAA06451.1; -.
 CC EMBL: AB032202; BAA92435.1; -.
 CC PDB: 1NPM; 23-WAR-99.
 CC MEROPS: S01.244; -.
 CC MGD: MGI:892018; Klk8.
 CC InterPro: IPR001314; Chymotrypsin.
 CC DR InterPro: IPR001254; Trypsin.
 CC Pfam: PF00089; trypsin; 1.
 CC PRINTS: PR00722; CHYMOTRYPSIN.
 CC SMART: SM00020; Tryp_Spc; 1.
 CC PROSITE: PS0240; TRYPSIN_DOM; 1.
 CC PROSITE: PS00134; TRYPSIN_HIS; 1.
 CC PROSITE: PS00135; TRYPSIN_SER; 1.
 CC Hydrolase; Serine protease; Zymogen; Signal;
 KW 3D-structure. 1 28 POTENTIAL.
 FT SIGNAL 29 32
 FT PROPEP 33 260
 FT CHAIN 33 260
 FT ACT_SITE 73 73
 FT ACT_SITE 120 120
 FT ACT_SITE 212 212
 FT ACT_SITE 39 173
 FT DISULFID 58 74
 FT DISULFID 145 246
 FT DISULFID 152 218
 FT DISULFID 184 198
 FT DISULFID 208 233
 FT CARBOHYD 110 110
 SQ SEQUENCE 260 AA; 28523 MW; BE5F6F6BE37CD60E CRC64;
 N-LINKED (GLCNAC...) (POTENTIAL).
 N-SEQUENCE 260 AA; 28523 MW; BE5F6F6BE37CD60E CRC64;
 Query Match 80.9%; Score 38; DB 1; Length 260;
 Best Local Similarity 77.8%; Pred. No. 4.5;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 VLVNERWVL 9
 ||||:|
 DB 61 VLVGDRWVL 69

```
RESULT 4
NRPN_RAT
ID NRPN_RAT STANDARD; PRT; 260 AA.
AC O88780;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neutropsin precursor (EC 3.4.21.-) (NP) (Kallikrein 8) (Brain serine
DE protease 1).
GN KLB OR PRSS19 OR NRPN OR BSPL.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=FISCHER; TISSUE=Brain;
RX MEDLINE=98389725; PubMed=9722524;
RA Davies B.J., Pickard B.S., Steel M., Morris R.G.M., Lathe R.;
RT "Serine proteases in rodent hippocampus.";
RL J. Biol. Chem. 273:23004-23011(1998).
CC -!- FUNCTION: SUGGESTED TO BE INVOLVED IN KINDLING EPILEPTOGENESIS AND
CC HIPPOCAMPAL PLASTICITY. HAS A STRONG PROTEOLYTIC ACTIVITY AGAINST
CC FIBRONECTIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- TISSUE SPECIFICITY: RESTRICTED TO HIPPOCAMPUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ005641; CAA06643.1; -.
DR HSSP; Q61955; INPM.
DR InterPro; IPR001314; Chymotrypsin.
DR ProSite; PS001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 32 BY SIMILARITY.
FT CHAIN 33 260 NEUROPSIN.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 39 173 BY SIMILARITY.
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 145 246 BY SIMILARITY.
FT DISULFID 152 218 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 208 233 BY SIMILARITY.
FT CARBOHYD 110 110 N-LINKED (GLCNAC...-) (POTENTIAL).
SQ SEQUENCE 260 AA; 28510 MW; 58DF4F0602A087F5 CRC64;
Query Match 80.9%; Score 38; DB 1; Length 260;
Best Local Similarity 77.8%; Pred. No. 4.5;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 VLVNRRWL 9
DB 61 VLVGDRWL 69
```

```
RESULT 5
TRYP_PLEPL
ID TRYP_PLEPL STANDARD; PRT; 250 AA.
AC P35034;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Trypsin precursor (EC 3.4.21.4).
OS Pleuronectes platessa (Plaice).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
OC Pleuronectoidei; Pleuronectidae; Pleuronectes.
OX NCBI_TaxID=8262;
[1]
RN SEQUENCE FROM N.A.
RP TISSUE=Liver;
RA Leaver M.J., George S.G.;
RL Submitted (NOV-1990) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X56744; CAA40068.1; -.
DR PIR; S31384; S31384.
DR HSSP; P00761; LEPT.
DR MEROPS; S01.151; -.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT PROPEP 16 22 ACTIVATION PEPTIDE.
FT CHAIN 23 250 TRYPSIN.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 106 106 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 29 163 BY SIMILARITY.
FT DISULFID 47 63 BY SIMILARITY.
FT DISULFID 133 236 BY SIMILARITY.
FT DISULFID 140 209 BY SIMILARITY.
FT DISULFID 174 188 BY SIMILARITY.
FT DISULFID 199 223 BY SIMILARITY.
FT SITE 197 197 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 250 AA; 27527 MW; 637DE96185CIABRA CRC64;
Query Match 78.7%; Score 37; DB 1; Length 250;
Best Local Similarity 66.7%; Pred. No. 6.6;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 VLVNRRWL 9
DB 50 VLVNRRWL 58
```

```
RESULT 6
VSP4_TRIMU
ID VSP4_TRIMU STANDARD; PRT; 257 AA.
AC Q91510;
```

```

DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucrofibrase 4 precursor (EC 3.4.21.-).
OS Trimeresurus mucrosquamatus (Taiwan habu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Protobothrops.
OX NCBI_TaxID=103944;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=95110313; PubMed=7811255;
RA Hung C.-C., Huang K.F., Chiou S.-H.;
RT "Characterization of one novel venom protease with beta-fibrinogenase
RT purification and cDNA sequence analysis.";
RL Biochem. Biophys. Res. Commun. 205:1707-1715(1994).
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. CLEAVES BETA-
CC CHAIN OF FIBRINOGEN MOLECULES EFFICIENTLY AND SHOWS RELATIVELY
CC LOWER ACTIVITY ON ALPHA-CHAIN, WITH ALMOST NO ACTIVITY ON GAMMA-
CC CHAIN.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83224; CAA58224.1; -
DR HSSP; P20231; IAAO.
DR MEROPS; S01.185; -.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN; 1.
DR PROSITE; PS00134; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; venom; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 257
FT ACT_SITE 64 64
FT ACT_SITE 109 109
FT ACT_SITE 203 203
FT DISULFID 31 162
FT DISULFID 49 65
FT DISULFID 97 255
FT DISULFID 141 209
FT DISULFID 173 188
FT DISULFID 199 224
FT CARBOHYD 102 102
SQ SEQUENCE 257 AA; 28284 MW; 86060B68CE6FCA8 CRC64;

Query Match 78.7%; Score 37; DB 1; Length 257;
Best Local Similarity 75.0%; Pred. No. 6.8;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 53 LINEEWVL 60

RESULT 7
VSP5_TRIMU STANDARD; PRT; 257 AA.

```

```

AC Q91511;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucrofibrase 5 precursor (EC 3.4.21.-).
OS Trimeresurus mucrosquamatus (Taiwan habu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Protobothrops.
OX NCBI_TaxID=103944;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=95110313; PubMed=7811255;
RA Hung C.-C., Huang K.F., Chiou S.-H.;
RT "Characterization of one novel venom protease with beta-fibrinogenase
RT purification and cDNA sequence analysis.";
RL Biochem. Biophys. Res. Commun. 205:1707-1715(1994).
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. CLEAVES BETA-
CC CHAIN OF FIBRINOGEN MOLECULES EFFICIENTLY AND SHOWS RELATIVELY
CC LOWER ACTIVITY ON ALPHA-CHAIN, WITH ALMOST NO ACTIVITY ON GAMMA-
CC CHAIN.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83225; CAA58225.1; -
DR HSSP; P20231; IAAO.
DR MEROPS; S01.185; -.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN; 1.
DR PROSITE; PS00134; TRYPSIN_DOM; 1.
DR PROSITE; PS00135; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Glycoprotein; venom; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 257
FT ACT_SITE 64 64
FT ACT_SITE 109 109
FT ACT_SITE 203 203
FT DISULFID 31 162
FT DISULFID 49 65
FT DISULFID 97 255
FT DISULFID 141 209
FT DISULFID 173 188
FT DISULFID 199 224
FT CARBOHYD 102 102
SQ SEQUENCE 257 AA; 28164 MW; FCC64A8EAF2827D CRC64;

Query Match 78.7%; Score 37; DB 1; Length 257;
Best Local Similarity 75.0%; Pred. No. 6.8;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 53 LINEEWVL 60

RESULT 8

```

VSP7_TRIMU
ID VSP7_TRIMU STANDARD; PRT; 257 AA.
AC Q9DG84;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serpentokallikrein-2 precursor (EC 3.4.21.-).
OS Trimeresurus mucrosquamatus (Taiwan habu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Metazoa; Chordata; Craniata; Vertebrata; Colubroidea;
OC Viperidae; Crotalinae; Scleroglossa; Serpentes; Colubroidea;
OX NCBI_TaxID=103944;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Chou S.-H., Hung C.-C.;
RT "Serpentokallikreins from Taiwan habu.";
RL Submitted (Oct-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF098262; AAG27253.1; -;
DR InterPro: IPR001314; Chymotrypsin.
DR Pfam: PF00089; Trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Venom; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 257 SERPENTOKALLIKREIN-2.
FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 162 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 97 255 BY SIMILARITY.
FT DISULFID 141 209 BY SIMILARITY.
FT DISULFID 173 188 BY SIMILARITY.
FT DISULFID 199 224 BY SIMILARITY.
FT CARBOHYD 102 102 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 257 AA; 28320 MW; 61EF5A1DC58F1E9 CRC64;

Query Match 78.7%; Score 37; DB 1; Length 257;
Best Local Similarity 75.0%; Pred. No. 6.8;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 53 LINEEWVL 60

RESULT 9
VSPC-TRIGA
ID VSPC-TRIGA STANDARD; PRT; 257 AA.
AC Q13062;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE VSP7_TRIMU STANDARD; PRT; 257 AA.
OS Trimeresurus gramineus (Indian green tree viper) (Green habu snake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidodonta; Metazoa; Chordata; Craniata; Vertebrata; Colubroidea;
OC Viperidae; Crotalinae; Scleroglossa; Serpentes; Colubroidea;
OX NCBI_TaxID=8767;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA MEDLINE=97096898; PubMed=8941719;
RA Deshimaru M., Ogawa T., Nakashima K., Nobuhisa I., Chijiwa T.,
RA Shimohigashi Y., Fukumaki Y., Niwa M., Yamashina I., Hattori S.,
RA Ohno M.;
RT "Accelerated evolution of crotalinae snake venom gland serine
RT proteases.";
RL FEBS Lett. 397:83-88(1996).
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D67084; BAA19982.1; -;
DR HSP: P00763; IDPO.
DR MEROPS: S01.185; -;
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Venom; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 257 VENOM SERINE PROTEINASE 2C.
FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 162 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 97 255 BY SIMILARITY.
FT DISULFID 141 209 BY SIMILARITY.
FT DISULFID 173 188 BY SIMILARITY.
FT DISULFID 199 224 BY SIMILARITY.
FT CARBOHYD 116 116 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 120 120 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 121 121 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 257 AA; 28042 MW; 2F50B6947CB33AB1 CRC64;

Query Match 78.7%; Score 37; DB 1; Length 257;
Best Local Similarity 75.0%; Pred. No. 6.8;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 53 LINEEWVL 60

RESULT 10
CTR2-ANOGA
ID CTR2-ANOGA STANDARD; PRT; 258 AA.
AC Q17025; Q17026;

```
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chymotrypsin 2 precursor (EC 3.4.21.1).
GN CHYM2.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SUAKOKO; TISSUE=Midgut;
RA Mueller H.M.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SUAKOKO; TISSUE=Midgut;
RA Mueller H.M.; Catteruccia F.; Crisanti A.;
RT "An Anopheles gambiae locus containing the sequences of two closely
RT related chymotrypsin-like proteases induced in the gut following blood
RT meal.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Trp-I-Xaa,
CC Phe-I-Xaa, Leu-I-Xaa.
CC -!- TISSUE SPECIFICITY: MIDGUT.
CC -!- DEVELOPMENTAL STAGE: INDUCED IN THE MIDGUT OF FEMALE AFTER BLOOD
CC MEAL.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z18888; CAA79326.1; -
DR EMBL; Z32645; CAA83567.1; -
DR HSSP; P00734; 2HNT.
DR MEROPS; S01.121; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal; Zymogen; Digestion;
KW Multigene family.
FT SIGNAL 1 17 POTENTIAL.
FT PROPEP 18 32 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 33 258 CHYMOTRYPSIN 2.
FT ACT_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 119 119 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 59 75 BY SIMILARITY.
FT DISULFID 182 198 BY SIMILARITY.
FT DISULFID 208 232 BY SIMILARITY.
FT SITE 206 206 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
FT CONFLICT 166 166 P -> R (IN REF. 1).
FT CONFLICT 194 196 LGH -> FPD (IN REF. 1).
SQ SEQUENCE 258 AA; 27919 MW; A41AD20A630BC67B CRC64;

Query Match 78.7%; Score 37; DB 1; Length 258;
Best Local Similarity 75.0%; Pred No. 6.9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
1:|:||||
```

```
Db 63 LLNDRWVL 70

RESULT 11
CTRL_ANOGA
ID CTRL_ANOGA STANDARD; PRT; 259 AA.
AC Q27289;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Chymotrypsin 1 precursor (EC 3.4.21.1).
GN CHYM1.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SUAKOKO; TISSUE=Midgut;
RA Mueller H.M.;
RL Submitted (NOV-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SUAKOKO; TISSUE=Midgut;
RA Mueller H.M.; Catteruccia F.; Crisanti A.;
RT "An Anopheles gambiae locus containing the sequences of two closely
RT related chymotrypsin-like proteases induced in the gut following blood
RT meal.";
RL Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa, Trp-I-Xaa,
CC Phe-I-Xaa, Leu-I-Xaa.
CC -!- TISSUE SPECIFICITY: MIDGUT.
CC -!- DEVELOPMENTAL STAGE: INDUCED IN THE MIDGUT OF FEMALE AFTER BLOOD
CC MEAL.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z18887; CAA79325.1; -
DR EMBL; Z32645; CAA83568.1; -
DR HSSP; P00766; 1CHG.
DR MEROPS; S01.121; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal; Zymogen; Digestion;
KW Multigene family.
FT SIGNAL 1 17 POTENTIAL.
FT PROPEP 18 32 ACTIVATION PEPTIDE.
FT CHAIN 33 259 CHYMOTRYPSIN 1.
FT ACT_SITE 74 74 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 119 119 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 59 75 BY SIMILARITY.
FT DISULFID 182 198 BY SIMILARITY.
FT DISULFID 208 232 BY SIMILARITY.
FT SITE 206 206 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 259 AA; 27717 MW; 66A9C0BA9D1E8C88 CRC64;

Query Match 78.7%; Score 37; DB 1; Length 259;
```

Best Local Similarity 75.0%; Pred. No. 6.9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I:|:||||
Db 63 LVNRWVL 70

RESULT 12
VSP1_TRIJE STANDARD; PRT; 260 AA.
AC Q9DF68;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Venom serine proteinase 1 precursor (EC 3.4.21.-) (SPL).
OS Trimeresurus jerdoni.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Trimeresurus.
OX NCBI_TaxID=135726;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Lu Q., Jin Y., Wei J., Wang W., Xiong Y.;
RT "cDNA cloning of serine proteinases from the venom of Trimeresurus jerdoni";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

EMBL; AF292110; AAG10788.1; -
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE: PS00135; TRYPSIN_SER; FALSE_NEG.
KW Hydrolase; Serine protease; Glycoprotein; Venom; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 24 BY SIMILARITY.
FT CHAIN 25 260 VENOM SERINE PROTEINASE 1.
FT ACT_SITE 67 67 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 206 206 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 165 BY SIMILARITY.
FT DISULFID 52 68 BY SIMILARITY.
FT DISULFID 100 258 BY SIMILARITY.
FT DISULFID 144 212 BY SIMILARITY.
FT DISULFID 176 191 BY SIMILARITY.
FT DISULFID 202 227 BY SIMILARITY.
FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 260 AA; 28614 MW; 1692283141E12896 CRC64;

Query Match 78.7%; Score 37; DB 1; Length 260;
Best Local Similarity 75.0%; Pred. No. 6.9;
Matches 6; Conservative 1; Mismatches 1; Indels 1; Gaps 0;

QY 2 LVNERWVL 9
I:|:||||
Db 56 LINEEWVL 63

RESULT 13
VSP2_TRIFL STANDARD; PRT; 260 AA.
AC Q13057;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Venom serine proteinase 2 precursor (EC 3.4.21.-).
GN TLF2.
OS Trimeresurus flavoviridis (Habu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Trimeresurus.
OX NCBI_TaxID=88087;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=97096898; PubMed=8941719;
RA Deshimaru M., Ogawa T., Nakashima K., Nobuhisa I., Chijiwa T.,
RA Shimohigashi Y., Fukumaki Y., Niwa M., Yamashina I., Hattori S.,
RA Ohno M.;
RT "Accelerated evolution of crotalinae snake venom gland serine proteases";
RL FEBS Lett. 397:83-88(1996).
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
EMBL; D67079; BAA19977.1; -
DR HSPSP; P00760; IAQ7.
DR MEROPS: S01.185; -
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS00240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE: PS00135; TRYPSIN_SER; FALSE_NEG.
KW Hydrolase; Serine protease; Glycoprotein; Venom; Zymogen; Signal.
FT SIGNAL 1 18 BY SIMILARITY.
FT PROPEP 19 24 BY SIMILARITY.
FT CHAIN 25 260 VENOM SERINE PROTEINASE 2.
FT ACT_SITE 67 67 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 206 206 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 165 BY SIMILARITY.
FT DISULFID 52 68 BY SIMILARITY.
FT DISULFID 100 258 BY SIMILARITY.
FT DISULFID 144 212 BY SIMILARITY.
FT DISULFID 176 191 BY SIMILARITY.
FT DISULFID 202 227 BY SIMILARITY.
FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 124 124 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 260 AA; 28641 MW; D0D0A5394CBF9B4A CRC64;

Query Match 78.7%; Score 37; DB 1; Length 260;
Best Local Similarity 55.6%; Pred. No. 6.9;

Matches 5: Conservative 3: Mismatches 1: Indels 0: Gaps 0;

QY 1 VLVNERWVL 9
: : : : :
Db 55 ILLINQEWVL 63

RESULT 14
KLKC_HUMAN
ID KLKC_HUMAN STANDARD; PRT; 248 AA.
AC Q9UKR0: Q9UKR1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kallikrein 12 precursor (EC 3.4.21.-) (kallikrein-like protein 5)
GN (KLK-L5).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20118156; PubMed=10652563;
RA Yousef G.M., Luo L.-Y., Diamandis E.P.;
RT "Identification of novel human kallikrein-like genes on chromosome
19q13.3-q13.4";
RL Anticancer Res. 19:2843-2852(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA Yousef G.M., Magklara A., Scorillas A., Diamandis E.P.;
RT "Cloning of new alternatively spliced forms of the kallikrein-like
gene 5 (KLK-L5)";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20510030; PubMed=11054574;
RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
RA Moss P., Paepel B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene
cluster located in chromosome 19q13 region.";
RL Gene 257:119-130(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhart-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Garnes J.,
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Andreise T., Frankheim M., Attix C., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of chromosome 19q13.4";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC produced by alternative splicing.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AF135025; RAD26426.2;
DR EMBL; AF135025; AAF06065.1;
DR EMBL; AF243527; AAG33365.1;
DR EMBL; AC011473; AAG23258.1;
DR HSP; P00761; 1AKS.

DR MEROPS; S01.020; -.
DR MIM; 605539; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 17
FT CHAIN 18 248
FT ACT_SITE 62 62
FT ACT_SITE 108 108
FT ACT_SITE 200 200
FT DISULFID 28 161
FT DISULFID 47 63
FT DISULFID 133 235
FT DISULFID 140 206
FT DISULFID 172 186
FT DISULFID 196 222
FT CARBOHYD 24 24
FT CARBOHYD 163 163
FT VARSPLIC 236 248
SQ SEQUENCE 248 AA; 26733 MW; BB473E98F8BAF703 CRC64;
Query Match 76.6%; Score 36; DB 1; Length 248;
Best Local Similarity 66.7%; Pred. No. 10;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
: : : : :
Db 50 VLIDHRWVL 58

RESULT 15
GRAM_RAT
ID GRAM_RAT STANDARD; PRT; 258 AA.
AC Q03238;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Granzyme M precursor (EC 3.4.21.-) (MET-ASE) (Natural killer cell
granular protease) (NK-MET-1) (Fragment).
GN GZMM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 21-44.
RX MEDLINE=93077530; PubMed=1447189;
RA Smyth M.J., Wiltout T., Trapani J.A., Ottaway K.S., Sowder R.,
RA Henderson L.E., Kam C.M., Powers J.C., Young H.A., Sayers T.J.;
RT "Purification and cloning of a novel serine protease, RNK-Met-1, from
the granules of a rat natural killer cell leukemia";
RL J. Biol. Chem. 267:24418-24425(1992).
CC -!- FUNCTION: CLEAVES PEPTIDE SUBSTRATES AFTER METHIONINE, LEUCINE,
CC AND NORLEUCINE.
CC -!- SUBCELLULAR LOCATION: GRANULES OF LARGE GRANULAR LYMPHOCYTES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. GRANZYME SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

```
-----
CC EMBL; L05175; AAA42056.1;
DR PIR; A45161; A45161.
DR HSSP; 061955; 1NPM.
DR MEROPS; S01.139;
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM0020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Signal; Zymogen; T-cell;
KW Cytolysis.
FT NON_TER 1 1
FT SIGNAL <1 2 POTENTIAL.
FT PROPEP 2 20 ACTIVATION PEPTIDE.
FT CHAIN 21 258 GRANZYME M.
FT ACT_SITE 61 61 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 107 107 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 204 204 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 46 62 BY SIMILARITY.
FT DISULFID 142 210 BY SIMILARITY.
FT DISULFID 173 189 BY SIMILARITY.
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 258 AA; 28339 MW; B89DC10EF54DF495 CRC64;

Query Match 76.6%; Score 36; DB 1; Length 258;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 VLVNERWVL 9
| | | | |
DB 49 VLVHQKWL 57

RESULT 16
ESTA_CANFA STANDARD; PRT; 260 AA.
AC P09582;
DT 01-WAR-1989 (Rel. 10, Created)
DT 01-WAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Arginine esterase precursor (EC 3.4.21.35).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=88211858; PubMed=2835268;
RA Chapdelaine P., Ho-Kim M.-A., Tremblay R.R., Dube J.Y.;
RT "Nucleotide sequence of the androgen-dependent arginine esterase mRNA
of canine prostate.";
RL FEBS Lett. 232:187-192(1988).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91119675; PubMed=1991049;
RA Chapdelaine P., Gauthier E., Ho-Kim M.A., Bissonnette L.,
Tremblay R.R., Dube J.Y.;
RT "Characterization and expression of the prostatic arginine esterase
gene, a canine glandular kallikrein.";
RL DNA Cell Biol. 10:49-59(1991).
RN [3]
RP SEQUENCE OF 25-50 AND 108-145.
RC TISSUE=Prostate;
RX MEDLINE=85004070; PubMed=6566614;
RA Lazure C., Leduc R., Seidah N.G., Chretien M., Dube J.Y.,
Chapdelaine P., Frenette G., Paquin R., Tremblay R.R.;
RT "The major androgen-dependent protease in dog prostate belongs to the
kallikrein family: confirmation by partial amino acid sequencing.";
RL FEBS Lett. 175:1-7(1984).

-----
CC EMBL; Y00751; CAA68720.1;
DR EMBL; M63669; AAA30831.1;
DR PIR; A30981; A30981.
DR PIR; A37938; A37938.
DR PIR; S00613; S00613.
DR HSSP; P00752; 2PKA.
DR MEROPS; S01.160;
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 17 POTENTIAL.
FT PROPEP 18 24 ACTIVATION PEPTIDE.
FT CHAIN 25 260 ARGININE ESTERASE.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 119 119 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (PROBABLE).
FT DISULFID 31 172 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 151 218 BY SIMILARITY.
FT DISULFID 183 197 BY SIMILARITY.
FT DISULFID 208 233 BY SIMILARITY.
FT CONFLICT 56 56 N -> H (IN REF. 2).
SQ SEQUENCE 260 AA; 28746 MW; 48768B6EF204775A CRC64;

Query Match 76.6%; Score 36; DB 1; Length 260;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 VLVNERWVL 9
| | | | |
DB 53 VLVNPEWVL 61

RESULT 17
VA20_VACCC STANDARD; PRT; 426 AA.
AC P20995;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein A20.
GN A20R;
OS Vaccinia virus (strain Copenhagen).
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus.
OX NCBI_TaxID=10249;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91021027; PubMed=2219722;
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
Paoletti E.;
RT "The complete DNA sequence of vaccinia virus.";
```

```

RL Virology 179:247-266(1990)).
RN [2]
RP COMPLETE GENOME.
RA Goebel S.J., Johnson G.P., Perkus M.E., Davis S.W., Winslow J.P.,
RT Paolletti E.; 'The complete DNA sequence of vaccinia virus'.
RL Virology 179:517-563(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35027; AAA48143.1; -
DR PIR; D42519; D42519.
DR PIR; D42519; D42519.
SQ SEQUENCE 426 AA; 49187 MW; F316F8FDEC861DE3 CRC64;
Query Match 76.6%; Score 36; DB 1; Length 426;
Best Local Similarity 55.6%; Pred. No. 18;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 VLVNERWVL 9
DB 130 VLLNNRWIM 138
RESULT 18
VA20_VARV STANDARD; PRT; 426 AA.
AC P33843;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE Protein A20.
GN A20R OR A21R.
OS Variola virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Orthopoxvirus
OX NCBI_TaxID=10255;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-INDIA-1967 / ISOLATE IND3;
RX MEDLINE=93202281; PubMed=8384129;
RA Shchelkunov S.N., Blinov V.M., Sandakhchiev L.S.;
RT "Genes of variola and vaccinia viruses necessary to overcome the host
RT protective mechanisms.";
RL FEBS Lett. 319:80-83(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BANGLADESH-1975;
RX MEDLINE=94088747; PubMed=8264798;
RA Massung R.F., Esposito J.J., Liu L., Qi J., Utterback T.R.,
RA Knight J.C., Aubin L., Yuran T.E., Parsons J.M., Loparev V.N.,
RA Selivanov N.A., Cavallaro K.F., Kerlavage A.R., Mahy B.W.J.,
RA Venter C.J.;
RT "Potential virulence determinants in terminal regions of variola
RT smallpox virus genome.";
RL Nature 366:748-751(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X69198; CAA49065.1; -
DR EMBL; L22579; AAA60873.1; -
DR PIR; C36850; C36850.

```

```

SQ SEQUENCE 426 AA; 49088 MW; 6CD83C0903EBBF59 CRC64;
Query Match 76.6%; Score 36; DB 1; Length 426;
Best Local Similarity 55.6%; Pred. No. 18;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 VLVNERWVL 9
DB 130 VLLNNRWIM 138
RESULT 19
Y4BL_RHISN STANDARD; PRT; 516 AA.
AC P55379;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Putative transposase Y4BL/Y4KJ/Y4TB.
GN Y4BL AND Y4KJ AND Y4TB.
OS Rhizobium sp. (strain NGR234).
OG Plasmid sym pNGR234a.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -!- SIMILARITY: BELONGS TO THE IS21/IS408/IS1162 FAMILY OF
CC TRANSPOSASES.
CC -!- SIMILARITY: STRONG, TO Y4UI.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000066; AAB91627.1; -
DR EMBL; AE000081; AAB91741.1; -
DR EMBL; AE000097; AAB91856.1; -
DR InterPro; IPR001584; Rve.
DR Pfam; PF00665; rve; 1.
KW Hypothetical protein; Transposable element; Transposition;
KW DNA-binding; DNA recombination; Plasmid.
SQ SEQUENCE 516 AA; 58266 MW; 0ADF071C6DEA4A27 CRC64;
Query Match 76.6%; Score 36; DB 1; Length 516;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 VLVNERWVL 9
DB 266 VLVVERWIL 274
RESULT 20
TRY2_SALSA STANDARD; PRT; 231 AA.
AC P35032;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Trypsin II precursor (EC 3.4.21.4) (Fragment).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei.

```

```
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=96035908; PubMed=7556223;
RA Male R., Lorens J.B., Smals A.O., Torrisen K.R.;
RT "Molecular cloning and characterization of anionic and cationic
RL variants of trypsin from Atlantic salmon.";
RN Eur. J. Biochem. 232:677-685(1995).
RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).
RA Smalas A.O., Hordvik A.;
RT "Structure determination and refinement of benzamidine-inhibited
CC trypsin from the North Atlantic salmon (Salmo salar) at 1.82-A
CC resolution.";
CC Acta Crystallogr. D 49:318-330(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).
RX MEDLINE=95148588; PubMed=7846025;
RA Smalas A.O., Heimstad E.S., Hordvik A., Willassen N.P., Male R.;
RT "Cold adaption of enzymes: structural comparison between salmon and
CC bovine trypsins.";
CC Proteins 20:149-166(1994).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X70073; CAA49678.1; -
DR PIR; S31778; S31778.
DR HSSP; P35031; 1BIT.
DR MEROPS; S01.258; -.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; TRYPSPC; 1.
DR PROSITE; PS0240; TRYPSPIN_DOM; 1.
DR PROSITE; PS00134; TRYPSPIN_HIS; 1.
DR PROSITE; PS00135; TRYPSPIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT NON_TER 1 1
FT SIGNAL <1 4
FT PROPEP 5 9
FT CHAIN 10 231
FT ACT_SITE 49 49
FT ACT_SITE 93 93
FT ACT_SITE 185 185
FT DISULFID 16 145
FT DISULFID 34 50
FT DISULFID 118 218
FT DISULFID 125 191
FT DISULFID 156 170
FT DISULFID 181 205
FT SITE 179 179
FT SITE 179 179
SQ SEQUENCE 231 AA; 24823 MW; C54A1CAFE74FAE18 CRC64;
Query Match 74.5%; Score 35; DB 1; Length 231;
Best Local Similarity 75.0%; Pred. No. 14;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
DB 38 LVNENWV 45
RESULT 21
TRY1_SALSA
ID TRY1_SALSA STANDARD; PRT; 242 AA.
AC P35031.
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
DE Trypsin I precursor (EC 3.4.21.4).
OS Salmo salar (Atlantic salmon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;
OC Protacanthopterygii; Salmoniformes; Salmonidae; Salmo.
OX NCBI_TaxID=8030;
```

```
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pancreas;
RX MEDLINE=96035908; PubMed=7556223;
RA Male R., Lorens J.B., Smals A.O., Torrisen K.R.;
RT "Molecular cloning and characterization of anionic and cationic
RL variants of trypsin from Atlantic salmon.";
RN Eur. J. Biochem. 232:677-685(1995).
RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).
RA Smalas A.O., Hordvik A.;
RT "Structure determination and refinement of benzamidine-inhibited
CC trypsin from the North Atlantic salmon (Salmo salar) at 1.82-A
CC resolution.";
CC Acta Crystallogr. D 49:318-330(1993).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (1.82 ANGSTROMS).
RX MEDLINE=95148588; PubMed=7846025;
RA Smalas A.O., Heimstad E.S., Hordvik A., Willassen N.P., Male R.;
RT "Cold adaption of enzymes: structural comparison between salmon and
CC bovine trypsins.";
CC Proteins 20:149-166(1994).
CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-I-Xaa, Lys-I-Xaa.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X70075; CAA49680.1; -
DR EMBL; X70071; CAA49676.1; -
DR EMBL; X70072; CAA49677.1; -
DR PIR; S31775; S31775.
DR PIR; S31776; S31776.
DR PIR; S31777; S31777.
DR PDB; 2TBS; 30-APR-94.
DR PDB; 1BIT; 01-NOV-94.
DR MEROPS; S01.151; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSPC; 1.
DR PROSITE; PS0240; TRYPSPIN_DOM; 1.
DR PROSITE; PS00134; TRYPSPIN_HIS; 1.
DR PROSITE; PS00135; TRYPSPIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family; 3D-structure.
FT SIGNAL 1 15
FT PROPEP 16 20
FT CHAIN 21 242
FT ACT_SITE 60 60
FT ACT_SITE 104 104
FT ACT_SITE 196 196
FT DISULFID 27 156
FT DISULFID 45 61
FT DISULFID 129 229
FT DISULFID 136 202
FT DISULFID 167 181
FT DISULFID 192 216
FT SITE 190 190
FT VARIANT 33 33
SQ SEQUENCE 242 AA; 25958 MW; 43F5642498067E5A CRC64;
Query Match 74.5%; Score 35; DB 1; Length 242;
Best Local Similarity 75.0%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

QY 2 LVNERWVL 9
 DB 49 LVNERWV 56

RESULT 22
 TRV1_HUMAN
 ID TRV1_HUMAN STANDARD: PRT: 247 AA
 AC P07477; Q92955; Q9HAN4; Q9HAN5; Q9HAN6; Q9HAN7;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE TRYPsin I precursor (EC 3.4.21.4) (Cationic trypsinogen).
 GN PRSS1 OR TRY1 OR TRP1 OR TRYPL.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-86221712; PubMed-3011602;
 RA Emi M., Nakamura Y., Ogawa M., Yamamoto T., Nishide T., Mori T.,
 RA Matsubara K.;
 RT "Cloning, characterization and nucleotide sequences of two CDNAS
 RT encoding human pancreatic trypsinogens.";
 RL Gene 41:305-310(1986).
 RN [2]
 RP SEQUENCE OF 16-43.
 RX MEDLINE-90091010; PubMed-2598466;
 RA Kimlund M., Russick C., Marks W.H., Borgstroem A.;
 RT "Immunoreactive anionic and cationic trypsin in human serum.";
 RL Clin. Chim. Acta 184:31-46(1989).
 RN [3]
 RP SEQUENCE OF 68-151 FROM N.A., AND VARIANT HPC HIS-122.
 RX MEDLINE-96438847; PubMed-8841182;
 RA Whitcomb D.C., Gorry M.C., Preston R.A., Furey W., Sossenheimer M.J.,
 RA Ulrich C.D., Martin S.P., Gates L.K. Jr., Amann S.T., Toskes P.P.,
 RA Little R., McGrath K., Uomo G., Post J.C., Ehrlich G.D.;
 RT "Hereditary pancreatitis is caused by a mutation in the cationic
 RT trypsinogen gene.";
 RL Nat. Genet. 14:141-145(1996).
 RN [4]
 RP SEQUENCE OF 15-67 FROM N.A., AND VARIANT HPC GLY-22.
 RX MEDLINE-20389982; PubMed-10930381;
 RA Teich N., Ockenga J., Hoffmeister A., Manns M., Mossner J., Keim V.;
 RT "Chronic pancreatitis associated with an activation peptide mutation
 RT that facilitates trypsin activation.";
 RL Gastroenterology 119:461-465(2000).
 RN [5]
 RP SEQUENCE OF 68-151 FROM N.A., AND VARIANTS HPC P-104; C-116 AND F-139.
 RX Teich N., Bauer N., Mossner J., Keim V.;
 RA Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
 RN [6]
 RP X-RAY CRYSTALLOGRAPHY (2.2 ANGSTROMS), AND PHOSPHORYLATION.
 RX MEDLINE-96266496; PubMed-8683601;
 RA Gaboriaud C., Serre L., Guy-Crotte O., Forest E.,
 RA Fontecilla-Camps J.-C.;
 RT "Crystal structure of human trypsin 1: unexpected phosphorylation of
 RT Tyr151.";
 RL J. Mol. Biol. 259:995-1010(1996).
 RN [7]
 RP VARIANTS HPC ILE-29 AND HIS-122.
 RX MEDLINE-97463797; PubMed-9322498;
 RA Gorry M.C., Gabalzedeh D., Furey W., Gates L.K. Jr., Preston R.A.,
 RA Aston C.E., Zhang Y., Ulrich C., Ehrlich G.D., Whitcomb D.C.;
 RT "Mutations in the cationic trypsinogen gene are associated with
 RT recurrent acute and chronic pancreatitis.";
 RL Gastroenterology 113:1063-1068(1997).
 RN [8]
 RP VARIANT HPC ILE-29.
 RX MEDLINE-98295575; PubMed-9633818;
 RA Teich N., Mossner J., Keim V.;

RT "Mutations of the cationic trypsinogen in hereditary pancreatitis.";
 RL Hum. Mutat. 12:39-43(1998).
 RN [9]
 RP VARIANTS HPC VAL-16 AND HIS-122.
 RX MEDLINE-99315544; PubMed-10381903;
 RA Witt H., Luck W., Becker M.;
 RT "A signal peptide cleavage site mutation in the cationic trypsinogen
 RT gene is strongly associated with chronic pancreatitis.";
 RL Gastroenterology 117:7-10(1999).
 RN [10]
 RP VARIANT HPC ARG-23.
 RX MEDLINE-99219545; PubMed-10204851;
 RA Ferec C., Ragues O., Salomon R., Roche C., Bernard J.P., Guillot M.,
 RA Quere I., Faure C., Mercier B., Audrezet M.P., Guillausseau P.J.,
 RA Dupont C., Munnich A., Bignon J.D., Le Bodic L.;
 RT "Mutations in the cationic trypsinogen gene and evidence for genetic
 RT heterogeneity in hereditary pancreatitis.";
 RL J. Med. Genet. 36:228-232(1999).
 RN [11]
 RP VARIANTS HPC THR-29 AND CYS-122.
 RX PubMed-11788572;
 RA Pfutzer R., Myers E., Applebaum-Shapiro S., Finch R., Ellis I.,
 RA Neoptolemos J., Kant J.A., Whitcomb D.C.;
 RT "Novel cationic trypsinogen (PRSS1) N29T and R122C mutations cause
 RT autosomal dominant hereditary pancreatitis.";
 RL Gut 50:271-272(2002).
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
 CC -!- SUBCELLULAR LOCATION: Extracellular.
 CC -!- MASS SPECTROMETRY: MW=24348; MW_ERR=2; METHOD=Electrospray;
 CC RANGE=24-247.
 CC -!- DISEASE: defects in PRSS1 are a cause of hereditary pancreatitis,
 CC (HPC or HP); also known as chronic pancreatitis (CP). HPC is an
 CC autosomal dominant disease characterized by the presence of
 CC calcuuli in pancreatic ducts. It causes severe abdominal pain
 CC attacks.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPsin FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M22612; AAA61231.1; -;
 DR EMBL; U70137; AAC50728.1; -;
 DR EMBL; AF314534; AAG30943.1; -;
 DR EMBL; AF315309; AAG30947.1; -;
 DR EMBL; AF315310; AAG30948.1; -;
 DR EMBL; AF315311; AAG30949.1; -;
 DR PIR; A25852; A25852.
 DR PDB; 1TRN; 03-JUN-95.
 DR PDB; 1FXV; 17-JUN-98.
 DR MEROPS; S01.151; -;
 DR MIM; 276000; -;
 DR MIM; 167800; -;
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS00240; TRYPsin_DOM; 1.
 DR PROSITE; PS00134; TRYPsin_HIS; 1.
 DR PROSITE; PS00135; TRYPsin_SER; 1.
 KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
 KW Multigene family; Phosphorylation; 3D-structure; Disease mutation.
 FT SIGNAL 1 15
 FT PROPEP 16 23 ACTIVATION PEPTIDE.
 FT CHAIN 24 247 TRYPsin I.
 FT ACT_SITE 63 63 CHARGE RELAY SYSTEM.
 FT ACT_SITE 107 107 CHARGE RELAY SYSTEM.


```
DR PIR; A40197; DBHU.
DR PDB; IDFP; 25-FEB-98.
DR PDB; IDST; 11-JUL-96.
DR PDB; IDSU; 17-AUG-96.
DR MEROPS; S01.191; .
DR MIM; 134350;
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP-SPC; 1.
DR PROSITE; PS00240; TRYP-SIN_DOM; 1.
DR PROSITE; PS00134; TRYP-SIN_HIS; 1.
DR PROSITE; PS00135; TRYP-SIN_SER; 1.
KW Complement alternate pathway; plasma; Hydrolase; Serine protease;
KW zymogen; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 25 ACTIVATION PEPTIDE (POTENTIAL).
FT CHAIN 26 253 COMPLEMENT FACTOR D.
FT ACT_SITE 66 66 CHARGE RELAY SYSTEM.
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM.
FT ACT_SITE 208 208 CHARGE RELAY SYSTEM.
FT DISULFID 51 67
FT DISULFID 148 214
FT DISULFID 179 195
FT DISULFID 204 229
FT CONFLICT 26 26
FT CONFLICT 35 35
FT CONFLICT 40 40
FT CONFLICT 49 49
FT CONFLICT 52 52
FT CONFLICT 59 59
FT CONFLICT 63 63
FT CONFLICT 73 73
FT CONFLICT 83 86
FT CONFLICT 83 84
FT CONFLICT 94 95
FT CONFLICT 96 96
FT CONFLICT 136 136
FT CONFLICT 178 191
FT CONFLICT 243 243
FT CONFLICT 250 250
FT CONFLICT 250 250
SQ SEQUENCE 253 AA; 27004 MW; BD553B70BD5C6AD CRC64;

Query Match 74.5%; Score 35; DB 1; Length 253;
Best Local Similarity 77.8%; Pred. No. 16;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
   ||| |||
Db 54 VLVAEQWVL 62

RESULT 24
VSP1_TRIMU
ID VSP1_TRIMU STANDARD; PRT; 257 AA.
AC Q91507;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucrofibrase 1 precursor (EC 3.4.21.-).
OS Trimeresurus mucrosquamatus (Taiwan habu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Protobothrops.
OX NCBI_TaxID=103944;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=95110313; PubMed=7811255;
RA Hung C.-C., Huang K.F., Chiou S.-H.;
RT "Characterization of one novel venom protease with beta-fibrinogenase
```

```
RT activity from the Taiwan habu (Trimeresurus mucrosquamatus):
RT purification and cDNA sequence analysis.";
RL Biochem. Biophys. Res. Commun. 205:1707-1715(1994).
CC !- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. CLEAVES BETA-
CC CHAIN OF FIBRINOGEN MOLECULES EFFICIENTLY AND SHOWS RELATIVELY
CC LOWER ACTIVITY ON ALPHA-CHAIN, WITH ALMOST NO ACTIVITY ON GAMMA-
CC CHAIN.
CC !- SUBCELLULAR LOCATION: Secreted.
CC !- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYP-SIN FAMILY. SNAKE VENOM SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X83221; CAA58221.1; .
DR HSSP; P20231; IAAO.
DR MEROPS; S01.185; .
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP-SPC; 1.
DR PROSITE; PS00240; TRYP-SIN_DOM; 1.
DR PROSITE; PS00134; TRYP-SIN_HIS; 1.
DR PROSITE; PS00135; TRYP-SIN_SER; 1.
KW Hydrolase; Serine protease; Venom; Zymogen; Signal.
FT SIGNAL 1 18 BY SIMILARITY.
FT PROPEP 19 24 BY SIMILARITY.
FT CHAIN 25 257 MUCROFIBRASE 1.
FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 162 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 97 235 BY SIMILARITY.
FT DISULFID 141 209 BY SIMILARITY.
FT DISULFID 173 188 BY SIMILARITY.
FT DISULFID 199 224 BY SIMILARITY.
SQ SEQUENCE 257 AA; 28099 MW; 63BF3F31FB641416 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 257;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
   ||| |||
Db 53 LLNERWVL 60

RESULT 25
VSP2_TRIMU
ID VSP2_TRIMU STANDARD; PRT; 257 AA.
AC Q91508; Q91500;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Mucrofibrase 2 precursor (EC 3.4.21.-) (Trimubin).
OS Trimeresurus mucrosquamatus (Taiwan habu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Protobothrops.
OX NCBI_TaxID=103944;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=95110313; PubMed=7811255;
RA Hung C.-C., Huang K.F., Chiou S.-H.;
```

```

RT "Characterization of one novel venom protease with beta-fibrinogenase
RT activity from the Taiwan habu (Trimeresurus mucrosquamatus)":
RT purification and cDNA sequence analysis.";
RL Biochem. Biophys. Res. Commun. 205:1707-1715(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA Yaw-Wen G., Tseng-Yuan C., Mei-Chih L., Hsi-Hu W., Hwan-Wun L.,
RA Shou-Hsian M.;
RT Cloning and sequencing of a trimucin cDNA from Trimeresurus
RT mucrosquamatus.";
RL Submitted (JUL-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. CLEAVES BETA-
CC CHAIN OF FIBRINOGEN MOLECULES EFFICIENTLY AND SHOWS RELATIVELY
CC LOWER ACTIVITY ON ALPHA-CHAIN, WITH ALMOST NO ACTIVITY ON GAMMA-
CC CHAIN
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83222; CA58222.1; -.
DR EMBL; U31417; AAB01070.1; -.
DR HSSP; P20231; 1A0A.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Venom; Zymogen; Signal.
FT SIGNAL 1 18 BY SIMILARITY.
FT PROPEP 19 24 BY SIMILARITY.
FT CHAIN 25 257 MUCROFIBRASE 2.
FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 162 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 97 255 BY SIMILARITY.
FT DISULFID 141 209 BY SIMILARITY.
FT DISULFID 173 188 BY SIMILARITY.
FT DISULFID 199 224 BY SIMILARITY.
FT CONFLICT 159 159 V -> I (IN REF. 2).
SQ SEQUENCE 257 AA; 28149 MW; 3B39FF9D3506D8A4 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 257;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 53 LLNEEWVL 60

RESULT 26
VSP3_AGKAC STANDARD; PRT; 257 AA.
AC Q918X0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Venom serine proteinase Dav-KN precursor (EC 3.4.21.-).

```

```

OS Agkistrodon acutus (Hundred-pace snake) (Deinagkistrodon acutus).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Deinagkistrodon.
OX NCBI_TaxID=36307;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RA MEDLINE-21092922; PubMed-11171091;
RA Wang Y.-M., Wang S.-R., Tsai I.-H.;
RT "Serine protease isoforms of Deinagkistrodon acutus venom: cloning,
RT sequencing and phylogenetic analysis.";
RL Biochem. J. 354:161-168(2001).
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF159059; AAF76379.1; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Venom; Zymogen; Signal.
FT SIGNAL 1 18 BY SIMILARITY.
FT PROPEP 19 24 BY SIMILARITY.
FT CHAIN 25 257 VENOM SERINE PROTEINASE DAV-KN.
FT ACT_SITE 64 64 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 31 162 BY SIMILARITY.
FT DISULFID 49 65 BY SIMILARITY.
FT DISULFID 97 255 BY SIMILARITY.
FT DISULFID 173 188 BY SIMILARITY.
FT DISULFID 199 224 BY SIMILARITY.
SQ SEQUENCE 257 AA; 28333 MW; EABFD36F64EA4FC0 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 257;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 53 LLNEEWVL 60

RESULT 27
VSP3_TRIMU STANDARD; PRT; 257 AA.
AC Q91509;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Mucrofibrase 3 precursor (EC 3.4.21.-).
OS Trimeresurus mucrosquamatus (Taiwan habu).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Protobothrops.
OX NCBI_TaxID=103944;
RN [1]

```



```
RP SEQUENCE FROM N.A.
RX TISSUE=Venom gland;
RC MEDLINE=95110313; PubMed=7811255;
RA Hung C.-C., Huang K.F., Chiou S.-H.;
RT "Characterization of one novel venom protease with beta-fibrinogenase
RT activity from the Taiwan habu (Trimeresurus mucrosquamatus);
RL Biochem. Biophys. Res. Commun. 205:1707-1715(1994).
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. CLEAVES BETA-
CC CHAIN OF FIBRINOGEN MOLECULES EFFICIENTLY AND SHOWS RELATIVELY
CC LOWER ACTIVITY ON ALPHA-CHAIN, WITH ALMOST NO ACTIVITY ON GAMMA-
CC CHAIN.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X83223; CAA58223.1; -.
DR HSP; P20231; IAAO.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Venom; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 257
FT ACT_SITE 64 64
FT ACT_SITE 109 109
FT ACT_SITE 203 203
FT DISULFID 31 162
FT DISULFID 49 65
FT DISULFID 97 255
FT DISULFID 141 209
FT DISULFID 173 188
FT DISULFID 199 224
FT DISULFID 257 257
SQ SEQUENCE 257 AA; 28159 MW; 781E8531FB641416 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 257;
Best Local Similarity 75.0%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
DB 53 LLINEWVL 60
I:|I|I|I|

RESULT 28
FA9_SHEEP
ID FA9_SHEEP STANDARD; PRT; 274 AA.
AC P16291.
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15; Last sequence update)
DT 15-JUL-1999 (Rel. 38; Last annotation update)
DE Coagulation factor IX (EC 3.4.21.22) (Christmas factor) (Fragment).
GN F9.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
```

```
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=90152675; PubMed=2303254;
RA Sarkar G., Koebel D.D., Sommer S.S.;
RT "Direct sequencing of the activation peptide and the catalytic domain
RT of the factor IX gene in six species.";
RL Genomics 6:133-143(1990).
CC -!- FUNCTION: FACTOR IX IS A VITAMIN K-DEPENDENT PLASMA PROTEIN THAT
CC PARTICIPATES IN THE INTRINSIC PATHWAY OF BLOOD COAGULATION BY
CC CONVERTING FACTOR X TO ITS ACTIVE FORM IN THE PRESENCE OF CA++
CC IONS, PHOSPHOLIPIDS, AND FACTOR VIIIa.
CC -!- CATALYTIC ACTIVITY: Hydrolyses one Arg-Ile bond in factor X to
CC form factor Xa.
CC -!- SUBUNIT: FACTOR IX IS ACTIVATED BY FACTOR XIA, WHICH EXCISES THE
CC ACTIVATION PEPTIDE AND THUS PRODUCES A MOLECULE CONSISTING OF 2
CC CHAINS HELD TOGETHER BY ONE OR MORE DISULFIDE BONDS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M26233; AAA31520.1; -.
DR HSP; P00761; IAKS.
DR MEROPS; S01.214; -.
DR InterPro; IPR001254; Trypsin.
DR InterPro; IPR000294; VitK_dep_GLA.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00011; GLU_CARBOXYLATION; PARTIAL.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Blood coagulation; Plasma; Serine protease; Calcium-binding;
KW Hydrolase; Glycoprotein.
FT NON_TER 1 1
FT ACT_SITE 89 89
FT ACT_SITE 137 137
FT ACT_SITE 233 233
FT ACT_SITE 25 25
FT CARBOHYD 35 35
FT CARBOHYD 40 40
FT CARBOHYD 128 128
FT NON_TER 274 274
SQ SEQUENCE 274 AA; 30595 MW; D3617FC3B1D33E9B CRC64;

Query Match 74.5%; Score 35; DB 1; Length 274;
Best Local Similarity 62.5%; Pred. No. 17;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
DB 78 IVNERWV 85
:|I|I|I|I|

RESULT 29
RS2_AQUAE
ID RS2_AQUAE STANDARD; PRT; 274 AA.
AC O67809;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39; Last sequence update)
DT 16-OCT-2001 (Rel. 40; Last annotation update)
DE 30S ribosomal protein S2.
GN RPSB OR AQ.2007.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
```

```

RP SEQUENCE FROM N.A.
RC STRAIN-VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Snead M.A., Keller M., Aujay M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RT aeolicus.";
RL Nature 392:353-358(1998).
CC -!- SIMILARITY: BELONGS TO THE S2P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000767; AAC07767.1; -
DR InterPro; IPR001865; Ribosomal_S2.
DR Pfam; PF00318; Ribosomal_S2_1.
DR PRINTS; PR00395; RIBOSOMAL_S2.
DR PROSITE; PS00962; RIBOSOMAL_S2_1; 1.
DR PROSITE; PS00963; RIBOSOMAL_S2_2; FALSE_NEG.
KW Ribosomal protein; Complete proteome.
SQ SEQUENCE 274 AA; 31447 MW; 20A2903D25C2A649 CRC64;

Query Match 74.5%; Score 35; DB 1; Length 274;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNERWV 8
DB 92 VNERWV 97
|||||
RESULT 30
TEST_MOUSE
ID TEST_MOUSE STANDARD; PRT; 324 AA.
AC Q9JH7; Q9DA14;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Testisin precursor (EC 3.4.21.-) (Trypsin 4).
GN PRSS21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=21153229; PubMed=11231276;
RA Scarmen A.L., Hooper J.D., Boucatt K.J., Sit M.-L., Webb G.C.,
RA Normyle J.F., Antalis T.M.;
RT "Organization and chromosomal localization of the murine Testisin gene
RT encoding a serine protease temporally expressed during
RT spermatogenesis.";
RL Eur. J. Biochem. 268:1250-1258(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Testis;
RX PubMed=11259427;
RA Wong G.-W., Li L., Madhusudhan M.S., Krillis S.A., Gurish M.F.,
RA Rothenberg M.E., Sali A., Stevens R.L.;
RT "Trypsin 4, a new member of the chromosome 17 family of mouse serine
RT proteases.";
RL J. Biol. Chem. 276:20648-20658(2001).
RN [3]
RP SEQUENCE OF 3-324 FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Testis;
RX MEDLINE=21085660; PubMed=11217851;

```

```

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staabli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Anon H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wyshak-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- FUNCTION: COULD REGULATE PROTEOLYTIC EVENTS ASSOCIATED WITH
CC TESTICULAR GERM CELL MATURATION.
CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -!- TISSUE SPECIFICITY: TESTIS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED IN POST-MEIOTIC TESTICULAR GERM
CC CELLS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC -!- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A STOP
CC CODON IN POSITION 315.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF304012; AAK29360.1; -
DR EMBL; AY005145; AAG02255.1; -
DR EMBL; AF176209; AAF64407.2; -
DR EMBL; AF226710; AAF64428.2; -
DR EMBL; AK006271; BAB24495.1; ALT_SEQ.
DR MGD; MGI:1916698; Prss21.
DR InterPro; IPR003314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Signal; GPI-anchor; Zymogen.
RN SIGNAL 1 21
FT PROPEP 22 54 POTENTIAL.
FT CHAIN 55 298 TESTISIN.
FT PROPEP 299 324 REMOVED IN MATURE FORM (POTENTIAL).
FT ACT_SITE 95 95 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 147 147 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 248 248 CHARGE RELAY SYSTEM (POTENTIAL).
FT DISULFID 46 167 POTENTIAL.
FT DISULFID 80 96 POTENTIAL.
FT DISULFID 181 254 POTENTIAL.
FT DISULFID 214 233 POTENTIAL.
FT DISULFID 244 272 POTENTIAL.
FT LIPID 298 298 GPI-ANCHOR (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 177 177 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 210 210 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 283 283 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 275 275 P -> H (IN REF. 3).

```

```
SQ SEQUENCE 324 AA; 36175 MW; 56DC59B84F3C3CD4 CRC64;
Query Match 74.5%; Score 35; DB 1; Length 324;
Best Local Similarity 75.0%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I:| | | | |
Db 84 LLNRRWVL 91

RESULT 31
HPTI_CANFA
ID HPTI_CANFA STANDARD; PRT; 329 AA.
AC P19006;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Haptoglobin alpha and beta chains.
GN HP.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE.
RX MEDLINE=93054403; PubMed=1429498;
RA Kumazaki T., Urushibara N., Ishii S.-I.;
RT "Amino acid sequence and disulfide-bridge location of canine
haptoglobin.";
RL J. Biochem. 112:11-19(1992).
RN [2]
RP SEQUENCE OF 85-124.
RX MEDLINE=77025019; PubMed=975782;
RA Kurosky A., Kim H.H., Touchstone B.;
RT "Comparative sequence analysis of the N-terminal region of rat,
rabbit, and dog haptoglobin beta-chains.";
RL Comp. Biochem. Physiol. 55B:453-459(1976).
CC -!- FUNCTION: HAPTOGLOBIN COMBINES WITH FREE PLASMA HEMOGLOBIN,
PREVENTING LOSS OF IRON THROUGH THE KIDNEYS AND PROTECTING THE
KIDNEYS FROM DAMAGE BY HEMOGLOBIN, WHILE MAKING THE HEMOGLOBIN
ACCESSIBLE TO DEGRADATIVE ENZYMES.
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER, THEN RELEASED
IN PLASMA.
CC -!- MISCELLANEOUS: THE HEMOGLOBIN BINDING SITE IS LOCATED IN THE BETA
CHAIN.
CC -!- SIMILARITY: THE BETA CHAIN IS CLEARLY RELATED TO SERINE PROTEASES,
BUT HAPTOGLOBIN HAS NO PROTEOLYTIC ACTIVITY. PROBABLY BECAUSE THE
POSITIONS CORRESPONDING TO THE PROTEOLYTIC ACTIVE SITE RESIDUES IN
PROTEASES ARE OCCUPIED BY DIFFERENT AMINO ACIDS IN HAPTOGLOBIN.
DR PIR; JX0223; JX0223.
DR PIR; JX0224; JX0224.
DR PIR; B26503; B26503.
DR HSP; P00734; 2HNT.
DR MEROPS; S01.972; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
KW Glycoprotein; Serine protease homolog; Plasma; Hemoglobin-binding;
Liver.
FT CHAIN 1 83 HAPTOGLOBIN ALPHA CHAIN.
FT PROPEP 84 84 REMOVED (IN AT LEAST 80% OF THE ALPHA
CHAINS).
FT FT 85 329 HAPTOGLOBIN BETA CHAIN.
FT CHAIN 85 329 HAPTOGLOBIN BETA CHAIN.
FT DOMAIN 85 329 SERINE PROTEASE.
FT DISULFID 34 68 INTERCHAIN (ALPHA-BETA).
FT DISULFID 72 189
FT DISULFID 232 263

SQ SEQUENCE 274 304
FT CARBOHYD 9 9 N-LINKED (GLCNAC. . .).
FT CARBOHYD 107 107 N-LINKED (GLCNAC. . .).
FT CARBOHYD 214 214 N-LINKED (GLCNAC. . .).
SQ SEQUENCE 329 AA; 36457 MW; 86E32CA3E9CC2C48 CRC64;
Query Match 74.5%; Score 35; DB 1; Length 329;
Best Local Similarity 62.5%; Pred. No. 21;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I:| | | | |
Db 114 LINEQWLL 121

RESULT 32
HPTI_HUMAN
ID HPTI_HUMAN STANDARD; PRT; 347 AA.
AC P00737;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Haptoglobin-1 precursor.
GN HP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84247319; PubMed=6330675;
RA Brune J.L., Yang F., Barnett D.R., Bowman B.H.;
RT "Evolution of haptoglobin: comparison of complementary DNA encoding
Hp alpha 1S and Hp alpha 2FS.";
RL Nucleic Acids Res. 12:4531-4538(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=84158941; PubMed=6546723;
RA van der Straten A., Herzog A., Cabezon T., Bollen A.;
RT "Characterization of human haptoglobin cDNAs coding for alpha 2FS
beta and alpha 1S beta variants.";
RL FEBS Lett. 168:103-107(1984).
RN [3]
RP SEQUENCE OF 19-101 AND 103-347, AND DISULFIDE BONDS.
RX MEDLINE=81013890; PubMed=6997877;
RA Kurosky A., Barnett D.R., Lee T.H., Touchstone B., Hay R.E.,
RA Arnott M.S., Bowman B.H., Fitch W.M.;
RT "Covalent structure of human haptoglobin: a serine protease homolog.";
Proc. Natl. Acad. Sci. U.S.A. 77:3388-3392(1980).
RN [4]
RP SEQUENCE OF 3-347 FROM N.A.
RX MEDLINE=85257429; PubMed=4018023;
RA Bensl G., Raugel G., Klefenz H., Cortese R.;
RT "Structure and expression of the human haptoglobin locus.";
EMBO J. 4:119-126(1985).
RN [5]
RP SEQUENCE OF 3-347 FROM N.A.
RX MEDLINE=99425270; PubMed=10493829;
RA Loftus B.J., Kim U.-J., Sneddon V.P., Kalush F., Brandon R.,
RA Fuhrmann J., Mason T., Crosby M.L., Barnstead M., Cronin L.,
RA Deslattes Mays A., Cao Y., Xu R.X., Kang H.-L., Mitchell S.,
RA Eichler E.E., Harris P.C., Venter J.C., Adams M.D.;
RT "Genome duplications and other features in 12 Mb of DNA sequence from
human chromosome 16p and 16q.";
Genomics 60:295-308(1999).
RN [6]
RP DISULFIDE BONDS (ALPHA CHAIN).
RX MEDLINE=73166302; PubMed=4573324;
RA Malchy B., Dixon G.H.;
RT "Studies on the interchain disulfides of human haptoglobins.";
Can. J. Biochem. 51:249-264(1973).
CC -!- FUNCTION: HAPTOGLOBIN COMBINES WITH FREE PLASMA HEMOGLOBIN,
PREVENTING LOSS OF IRON THROUGH THE KIDNEYS AND PROTECTING THE
```

CC KIDNEYS FROM DAMAGE BY HEMOGLOBIN, WHILE MAKING THE HEMOGLOBIN
CC ACCESSIBLE TO DEGRADATIVE ENZYMES.
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA-1 AND TWO BETA CHAINS.
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER, THEN RELEASED
CC IN PLASMA.
CC -!- POLYMORPHISM: THE TWO MAIN ALLELES OF HP ARE CALLED HP1F (FAST)
CC AND HP1S (SLOW). THE SEQUENCE SHOWN HERE IS THAT OF THE HP1S FORM.
CC -!- MISCELLANEOUS: THE HAPTOGLOBIN-HEMOGLOBIN COMPLEX IS DEGRADED IN
CC THE LIVER (WHERE HAPTOGLOBIN IS ALSO SYNTHESIZED). THE PARENCHYMAL
CC CELLS OF THE LIVER HAVE A RECEPTOR THAT RECOGNIZES THE
CC CONFORMATIONAL CHANGE IN HAPTOGLOBIN THAT RESULTS FROM ITS BINDING
CC TO HEMOGLOBIN.
CC -!- SIMILARITY: ALTHOUGH THE BETA CHAIN IS CLEARLY RELATED TO SERINE
CC PROTEASES, HAPTOGLOBIN HAS NO ENZYMATIC ACTIVITY. POSITIONS
CC CORRESPONDING TO THE PROTEOLYTIC ACTIVE SITE RESIDUES OF THE
CC PROTEASES ARE OCCUPIED BY DIFFERENT AMINO ACIDS IN HAPTOGLOBIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X00637; CAA25267.1; -
CC EMBL: K01763; AAA52684.1; -
CC EMBL: X01793; CAA25926.1; -
CC EMBL: X01786; CAA25926.1; JOINED.
CC EMBL: X02206; CAA25926.1; JOINED.
CC EMBL: X01789; CAA25926.1; JOINED.
CC EMBL: X01791; CAA25926.1; JOINED.
CC EMBL: AC004682; AAC27432.1; -
CC PIR: A25739; HPHUL.
CC HSSP: P00734; 2HNT.
CC MEROPS: S01.972; -
CC SWISS-2DPAGE; P00737; HUMAN.
CC HSC-2DPAGE; P00737; HUMAN.
CC Siena-2DPAGE; P00737; -
CC MIM: 140100; -
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00084; sushi; 1.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00032; CCP; 1.
CC SMART: SM00020; Tryp_SPC; 1.
CC PROSITE: PS0240; TRYPSIN_DOM; 1.
CC Glycoprotein; Serine protease homolog; Plasma; Hemoglobin-binding;
KW Acute phase; Liver; Signal; Polymorphism.
FT SIGNAL 1 18
FT CHAIN 19 101 HAPTOGLOBIN-1 ALPHA-1 CHAIN.
FT CHAIN 103 347 HAPTOGLOBIN-1 BETA CHAIN.
FT DOMAIN 103 347 SERINE PROTEASE.
FT DISULFID 33 33 INTERCHAIN (ALPHA-1-ALPHA-1).
FT DISULFID 52 86 INTERCHAIN (ALPHA-1-BETA).
FT DISULFID 90 207 INTERCHAIN (ALPHA-1-BETA).
FT DISULFID 250 281
FT DISULFID 292 322 N-LINKED (GLCNAC. . .).
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .).
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .).
FT CARBOHYD 182 182 NE -> DK (IN HP1F).
FT VARIANT 70 71 /FTIG-VAR_005294.
SQ SEQUENCE 347 AA; 38452 MW; 47EFE07D5FE15EB8 CRC64;

Query Match
Best Local Similarity 74.5%; Score 35; DB 1; Length 347;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 2 LVNERWVL 9
I:|:|:|:|
Db 132 LINEQWLL 139

RESULT 33
HPT_ATEGE
ID HPT_ATEGE STANDARD; PRT; 347 AA.
AC P50417;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Haptoglobin precursor.
GN HP.
OS Ateles geoffroyi (Black-handed spider monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Ateleinae; Ateles.
OX NCBI_TaxID=9509;
RN [1]
RP SEQUENCE FROM N.A.
RA Erickson L.M., Becker L., Maeda N.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: HAPTOGLOBIN COMBINES WITH FREE PLASMA HEMOGLOBIN,
CC PREVENTING LOSS OF IRON THROUGH THE KIDNEYS AND PROTECTING THE
CC KIDNEYS FROM DAMAGE BY HEMOGLOBIN, WHILE MAKING THE HEMOGLOBIN
CC ACCESSIBLE TO DEGRADATIVE ENZYMES (BY SIMILARITY).
CC -!- SUBUNIT: TETRAMER OF TWO ALPHA AND TWO BETA CHAINS (BY
CC SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Extracellular.
CC -!- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER, THEN RELEASED
CC IN PLASMA (BY SIMILARITY).
CC -!- SIMILARITY: ALTHOUGH THE BETA CHAIN IS CLEARLY RELATED TO SERINE
CC PROTEASES, HAPTOGLOBIN HAS NO ENZYMATIC ACTIVITY. POSITIONS
CC CORRESPONDING TO THE PROTEOLYTIC ACTIVE SITE RESIDUES OF THE
CC PROTEASES ARE OCCUPIED BY DIFFERENT AMINO ACIDS IN HAPTOGLOBIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U04852; AAA03727.1; -
CC HSSP: P00734; 2HNT.
CC MEROPS: S01.972; -
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR000436; Sushi_SCR_CCP.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00084; sushi; 1.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00032; CCP; 1.
CC SMART: SM00020; Tryp_SPC; 1.
CC PROSITE: PS0240; TRYPSIN_DOM; 1.
CC Glycoprotein; Serine protease homolog; Plasma; Hemoglobin-binding;
KW Acute phase; Liver; Signal.
FT SIGNAL 1 18 BY SIMILARITY.
FT CHAIN 19 347 HAPTOGLOBIN.
FT CHAIN 19 101 HAPTOGLOBIN ALPHA CHAIN.
FT CHAIN 103 347 HAPTOGLOBIN BETA CHAIN.
FT DOMAIN 103 347 SERINE PROTEASE.
FT DISULFID 33 33 INTERCHAIN (WITH ALPHA) (BY SIMILARITY).
FT DISULFID 52 86 BY SIMILARITY.
FT DISULFID 90 207 INTERCHAIN (ALPHA-BETA) (BY SIMILARITY).
FT DISULFID 250 281 BY SIMILARITY.
FT DISULFID 292 322 BY SIMILARITY.
FT CARBOHYD 125 125 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 152 152 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 347 AA; 38476 MW; 209A809E7C17C21F CRC64;

Query Match 74.5%; Score 35; DB 1; Length 347;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
|:|:|:|:|
DB 132 LINEQWLL 139

RESULT 34
HPT2_HUMAN STANDARD; PRT; 348 AA.
AC P00739;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Haptoglobin-related protein precursor.
GN HPR.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93122805; PubMed=1478675;
RA Erickson L.M., Kim H.S., Maeda N.;
RT "Junctions between genes in the haptoglobin gene cluster of
RT primates.";
RL Genomics 14:948-958(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85207676; PubMed=2987228;
RA Maeda N.;
RT "Nucleotide sequence of the haptoglobin and haptoglobin-related gene
RT pair. The haptoglobin-related gene contains a retrovirus-like
RT element.";
RL J. Biol. Chem. 260:6698-6709(1985).
CC -!- MISCELLANEOUS; ALTHOUGH IT IS NOT KNOWN WHETHER THIS PROTEIN IS
CC SYNTHESIZED, IT HAS ALL THE STRUCTURAL CHARACTERISTICS OF
CC FUNCTIONAL HAPTOGLOBIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M69197; AAA88079.1; -.
DR EMBL; K03431; AAA88081.1; -.
DR EMBL; M10935; AAA88081.1; JOINED.
DR PIR; A00919; HPHUR.
DR HSP; P00734; 2HNT.
DR MEROPS; S01.974; -.
DR MIM; 140210; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR000436; Sushi_SCR_CCP.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00084; sushi; 1.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS50240; TRYPSPIN_DOM; 1.
KW Plasma; Serine protease homolog; Hemoglobin-binding; Signal.
FT SIGNAL 1 19
FT CHAIN 20 348 HAPTOGLOBIN-RELATED PROTEIN.
FT DOMAIN 104 348 SERINE PROTEASE.
FT SEQUENCE 348 AA; 39007 MW; CF91FF352B8182FA CRC64;

Query Match 74.5%; Score 35; DB 1; Length 348;
Best Local Similarity 62.5%; Pred. No. 22;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
|:|:|:|:|
DB 133 LINEQWLL 140

RESULT 35
HPT2_HUMAN STANDARD; PRT; 406 AA.
AC P00738;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Haptoglobin-2 precursor.
GN HP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93122805; PubMed=1478675;
RA Erickson L.M., Kim H.S., Maeda N.;
RT "Junctions between genes in the haptoglobin gene cluster of
RT primates.";
RL Genomics 14:948-958(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85207676; PubMed=2987228;
RA Maeda N.;
RT "Nucleotide sequence of the haptoglobin and haptoglobin-related gene
RT pair. The haptoglobin-related gene contains a retrovirus-like
RT element.";
RL J. Biol. Chem. 260:6698-6709(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=83300066; PubMed=6310599;
RA Yang F., Brune J.L., Baldwin W.D., Barnett D.R., Bowman B.H.;
RT "Identification and characterization of human haptoglobin cDNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 80:5875-5879(1983).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=84158941; PubMed=6546723;
RA van der Straten A., Herzog A., Cabezon T., Bollen A.;
RT "Characterization of human haptoglobin cDNAs coding for alpha 2FS
RT beta and alpha 1S beta variants.";
RL FEBS Lett. 168:103-107(1984).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=84057759; PubMed=6688992;
RA van der Straten A., Herzog A., Jacobs P., Cabezon T., Bollen A.;
RT "Molecular cloning of human haptoglobin cDNA: evidence for a single
RT mRNA coding for alpha 2 and beta chains.";
RL EMBO J. 2:1003-1007(1983).
RN [6]
RP SEQUENCE OF 3-406 FROM N.A.
RX MEDLINE=83299252; PubMed=6310515;
RA Raugi G., Bensi G., Colantuoni V., Romano V., Santoro C.,
RA Costanzo F., Cortese R.;
RT "Sequence of human haptoglobin cDNA: evidence that the alpha and beta
RT subunits are coded by the same mRNA.";
RL Nucleic Acids Res. 11:5811-5819(1983).
RN [7]
RP REVISION TO ALPHA-2 CHAIN FROM COMPARISONS WITH ALPHA-1 CHAIN.
RX MEDLINE=81013890; PubMed=6997877;
RA Kurosky A., Barnett D.R., Lee T.-H., Touchstone B., Hay R.E.,
RA Arnott M.S., Bowman B.H., Fitch W.M.;
RT "Covalent structure of human haptoglobin: a serine protease homolog.";
RL Proc. Natl. Acad. Sci. U.S.A. 77:3388-3392(1980).
RN [8]
RP SEQUENCE OF 1-19 FROM N.A.
RX MEDLINE=86219709; PubMed=3519135;

RA van der Straten A., Falque J.-C., Loriau R., Bollen A., Cabezón T.;
 RT "Expression of cloned human haptoglobin and alpha 1-antitrypsin
 RT complementary DNAs in Saccharomyces cerevisiae.";
 RN DNA 5:129-136(1986).
 RN [9]
 RP DISULFIDE BONDS.
 RA MEDLINE=73166302; PubMed=4573324;
 RX Malchy B., Dixon G.H.;
 RT "Studies on the interchain disulfides of human haptoglobins.";
 CC Can. J. Biochem. 51:249-264(1973).
 CC -1- FUNCTION: HAPTOGLOBIN COMBINES WITH FREE PLASMA HEMOGLOBIN,
 CC PREVENTING LOSS OF IRON THROUGH THE KIDNEYS AND PROTECTING THE
 CC KIDNEYS FROM DAMAGE BY HEMOGLOBIN, WHILE MAKING THE HEMOGLOBIN
 CC ACCESSIBLE TO DEGRADATIVE ENZYMES.
 CC -1- SUBUNIT: TETRAMER OF TWO ALPHA-2 AND TWO BETA CHAINS.
 CC -1- SUBCELLULAR LOCATION: Extracellular.
 CC -1- TISSUE SPECIFICITY: SYNTHESIZED IN THE LIVER, THEN RELEASED
 CC IN PLASMA.
 CC -1- POLYMORPHISM: IF POSITION 71 IS LYS, THEN POSITION 130 IS GLU, AND
 CC VICE VERSA, IN THE F-S CHAIN SHOWN. A COMMON HAPTOGLOBIN ALLELE
 CC CODES FOR THIS MOLECULE WITH A PARTIALLY DUPLICATED ALPHA-2 CHAIN.
 CC -1- SIMILARITY: ALTHOUGH THE BETA CHAIN IS CLEARLY RELATED TO SERINE
 CC PROTEASES, HAPTOGLOBIN HAS NO ENZYMIC ACTIVITY. POSITIONS
 CC CORRESPONDING TO THE PROTEOLYTIC ACTIVE SITE RESIDUES OF THE
 CC PROTEASES ARE OCCUPIED BY DIFFERENT AMINO ACIDS IN HAPTOGLOBIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M69197; AAA88078.1; -;
 DR EMBL; M10935; AAA88080.1; -;
 DR EMBL; X00442; CAA25137.1; -;
 DR EMBL; X00422; AAA52687.1; -;
 DR EMBL; L29394; AAA52685.1; -;
 DR EMBL; M13192; -; NOT_ANNOTATED_CDS.
 DR EMBL; M12387; AAA52688.1; -;
 DR PIR; A00918; HPH02.
 DR PIR; B25739; B25739.
 DR HSSP; P00763; IDPO.
 DR MEROPS; S01.973; -;
 DR SWISS-2DPAGE; P00738; HUMAN.
 DR Siena-2DPAGE; P00738; -;
 DR MIM; 140100; -;
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR000436; Sushi_SCR_CCP.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00084; sushi; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00032; CCP; 2.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS00240; TRYPSTN_DOM; 1.
 KW Glycoprotein. Serine protease homolog; Plasma; Hemoglobin-binding;
 KW Repeat; Signal.
 FT SIGNAL 1 18
 FT CHAIN 19 160 HAPTOGLOBIN-2 ALPHA-2 CHAIN.
 FT CHAIN 162 406 HAPTOGLOBIN-2 BETA CHAIN.
 FT DOMAIN 162 406 SERINE_PROTEASE.
 FT DISULFID 33 33 INTERCHAIN (ALPHA-2-ALPHA-2).
 FT DISULFID 52 86
 FT DISULFID 92 92 INTERCHAIN (ALPHA-2-ALPHA-2).
 FT DISULFID 111 145
 FT DISULFID 149 266
 FT DISULFID 309 340
 FT DISULFID 351 381
 FT CARBOHYD 184 184
 FT CARBOHYD 207 207 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .).

FT CARBOHYD 241 241 N-LINKED (GLCNAC. . .).
 FT REPEAT 29 88
 FT REPEAT 89 146
 FT CONFLICT 70 70 D -> N (IN REF. 3).
 FT CONFLICT 90 90 D -> E (IN REF. 5).
 FT CONFLICT 239 239 H -> Q (IN REF. 6).
 FT CONFLICT 367 367 D -> N (IN REF. 6).
 SQ SEQUENCE 406 AA; A98B56B2B1BB891E CRC64;
 Query Match 74.5%; Score 35; DB 1; Length 406;
 Best Local Similarity 62.5%; Pred. No. 26;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LVNERWVL 9
 DB 191 LINEQWLL 198
 I:|:|:|:|
 RESULT 36
 FA9_BOVIN
 ID FA9_BOVIN STANDARD; PRT; 416 AA.
 AC P00741.
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Coagulation factor IX (EC 3.4.21.22) (Christmas factor).
 GN F9
 OS Bos taurus (Bovine).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 CC Bovidae; Bovinae; Bos.
 CC NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE
 RX MEDLINE=80056619; PubMed=291916;
 RA Katayama K., Ericsson L.H., Enfield D.L., Walsh K.A., Neurath H.,
 RA Davie E.W., Titani K.;
 RT "Comparison of amino acid sequence of bovine coagulation factor IX
 RT (Christmas Factor) with that of other vitamin K-dependent plasma
 RT proteins";
 RN Proc. Natl. Acad. Sci. U.S.A. 76:4990-4994(1979).
 RN [2]
 RP REVISION TO 64.
 RX MEDLINE=83308813; PubMed=6688526;
 RA McMullen B.A., Fujikawa K., Kisiel W.;
 RT "The occurrence of beta-hydroxyaspartic acid in the vitamin
 RT K-dependent blood coagulation zymogens";
 RN Biochem. Biophys. Res. Commun. 115:8-14(1983).
 RN [3]
 RP SEQUENCE OF 51-111 FROM N.A.
 RX MEDLINE=82272386; PubMed=6287289;
 RA Choo K.H., Gould K.G., Rees D.J.G., Brownlee G.G.;
 RT "Molecular cloning of the gene for human anti-haemophilic factor IX";
 RN Nature 299:178-180(1982).
 RN [4]
 RP STRUCTURE OF CARBOHYDRATE ON SER-53.
 RX MEDLINE=89213999; PubMed=3149637;
 RA Hase S., Kawabata S., Nishimura H., Takeya H., Sueyoshi T.,
 RA Miyata T., Iwanaga S., Takao T., Shimonishi Y., Ikenaka T.;
 RT "A new trisaccharide sugar chain linked to a serine residue in bovine
 RT blood coagulation factors VII and IX";
 RN J. Biochem. 104:867-868(1988).
 RN [5]
 RP STRUCTURE OF CARBOHYDRATE ON SER-53.
 RX MEDLINE=91344709; PubMed=2129367;
 RA Iwanaga S., Nishimura H., Kawabata S., Kisiel W., Hase S., Ikenaka T.;
 RT "A new trisaccharide sugar chain linked to a serine residue in the
 RT first EGF-like domain of clotting factors VII and IX and protein z.";
 RN Adv. Exp. Med. Biol. 281:121-131(1990).
 RN [6]
 RP STRUCTURE OF CARBOHYDRATE ON SER-53.
 RX MEDLINE=90130422; PubMed=2105311;
 RA Hase S., Nishimura H., Kawabata S., Iwanaga S., Ikenaka T.;

RT "The structure of (xylose)2glucose-O-serine 53 found in the first
 RT epidermal growth factor-like domain of bovine blood clotting factor
 RT IX.";
 RL J. Biol. Chem. 265:1858-1861(1990).
 CC -|- FUNCTION: FACTOR IX IS A VITAMIN K-DEPENDENT PLASMA PROTEIN THAT
 CC PARTICIPATES IN THE INTRINSIC PATHWAY OF BLOOD COAGULATION BY
 CC CONVERTING FACTOR X TO ITS ACTIVE FORM IN THE PRESENCE OF CA++
 CC IONS, PHOSPHOLIPIDS, AND FACTOR VIIIa.
 CC -|- CATALYTIC ACTIVITY: Hydrolyses one Arg-Ile bond in factor X to
 CC form factor Xa.
 CC -|- SUBUNIT: FACTOR IX IS ACTIVATED BY FACTOR XIA, WHICH EXCISES THE
 CC ACTIVATION PEPTIDE AND THUS PRODUCES A MOLECULE CONSISTING OF 2
 CC CHAINS HELD TOGETHER BY ONE OR MORE DISULFIDE BONDS.
 CC -|- MISCELLANEOUS: CALCIUM BINDS TO THE GAMMA-CARBOXYGLUTAMIC ACID
 CC (GLA) RESIDUES AND, WITH STRONGER AFFINITY, TO ANOTHER SITE,
 CC BEYOND THE GLA DOMAIN.
 CC -|- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J00007; AAA30520.1; -;
 DR PIR: A00923; KFBQ.
 DR HSP: P00740; ICFH.
 DR MEROPS: S01.214; -;
 DR GlycoSuiteDB: P00741; -;
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF.Ca.
 DR InterPro: IPR001438; EGF II.
 DR InterPro: IPR002383; GLA_blood.
 DR InterPro: IPR001254; Trypsin.
 DR InterPro: IPR000294; VitK_dep_GLA.
 DR Pfam: PF00008; EGF; 2.
 DR Pfam: PF00594; gla; 1.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR PRINTS: PR00010; EGFBL00D.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00001; EGF_Like; 1.
 DR SMART: SM00069; GLA; 1.
 DR SMART: SM00020; Tryp_Spc; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.
 DR PROSITE: PS00022; EGF_1; 1.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS00011; GLU_CARBOXYLATION; 1.
 DR PROSITE: PS00240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Blood coagulation; plasma; Serine protease; Calcium-binding;
 KW Hemophilias; Hydroxylation; Vitamin K; Gamma-carboxyglutamic acid;
 FT CHAIN 1 146 FACTOR IXA (ACTIVE FORM) LIGHT CHAIN.
 FT PROPEP 147 181 ACTIVATION PEPTIDE.
 FT CHAIN 182 416 FACTOR IXA (ACTIVE FORM) HEAVY CHAIN.
 FT DOMAIN 47 83 EGF-LIKE 1, CALCIUM-BINDING (POTENTIAL).
 FT DOMAIN 84 125 EGF-LIKE 2.
 FT DOMAIN 182 416 SERINE PROTEASE.
 FT SITE 146 147 CLEAVAGE (BY FACTOR XIA).
 FT SITE 181 182 CLEAVAGE (BY FACTOR XIA).
 FT MOD_RES 7 7 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 8 8 GAMMA-CARBOXYGLUTAMIC ACID.

FT MOD_RES 15 15 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 17 17 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 20 20 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 21 21 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 26 26 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 27 27 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 30 30 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 33 33 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 36 36 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 40 40 GAMMA-CARBOXYGLUTAMIC ACID.
 FT MOD_RES 64 64 HYDROXYLATION.
 FT DISULFID 18 23 BY SIMILARITY.
 FT DISULFID 51 62 BY SIMILARITY.
 FT DISULFID 56 71 BY SIMILARITY.
 FT DISULFID 73 82 BY SIMILARITY.
 FT DISULFID 88 99 BY SIMILARITY.
 FT DISULFID 95 109 BY SIMILARITY.
 FT DISULFID 111 124 BY SIMILARITY.
 FT CARBOHYD 53 53 O-LINKED (GLC...).
 FT CARBOHYD 158 158 N-LINKED (GLC...).
 FT CARBOHYD 158 168 N-LINKED (GLC...).
 FT CARBOHYD 173 173 N-LINKED (GLC...).
 FT CARBOHYD 261 261 N-LINKED (GLC...).
 FT ACT_SITE 222 222 CHARGE RELAY SYSTEM.
 FT ACT_SITE 270 270 CHARGE RELAY SYSTEM.
 FT ACT_SITE 366 366 CHARGE RELAY SYSTEM.
 FT VARIANT 64 64 D -> T (IN REF.1).
 SQ SEQUENCE 416 AA; 46785 MW; 34A/DFE916330662 CRC64;
 Query Match 74.5%; Score 35; DB 1; Length 416;
 Best Local Similarity 62.5%; Pred. No. 27;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LVNERWVL 9
 Db 211 IVNEKVVV 218
 :|||:|:
 RESULT 37
 ID ACRO_HUMAN STANDARD; PRT; 421 AA.
 AC P10323; Q9NU35;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Acrosin precursor (EC 3.4.21.10).
 GN AC OR ACRS.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Testis;
 RX MEDLINE=89153568; PubMed=2493394;
 RA Baba T., Watanabe K., Kashiwabara S.-I., Arai Y.;
 RT "Primary structure of human proacrosin deduced from its cDNA
 RT sequence.";
 RL FEBS Lett. 244:296-300(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Leukocyte;
 RX MEDLINE=90306003; PubMed=2114285;
 RA Kelme S., Adham I.M., Engel W.;
 RT "Nucleotide sequence and exon-intron organization of the human
 RT proacrosin gene.";
 RL Eur. J. Biochem. 190:195-200(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92331659; PubMed=1628652;
 RA Vazquez-Levin M.H., Reventos J., Gordon J.W.;
 RT "Molecular cloning, sequencing and restriction mapping of the genomic

sequence encoding human proacrosin.";
 Eur. J. Biochem. 207:23-26(1992).

RP DISCUSSION ON ABOVE PAPER.

RA Adham I.A., Spitzer U., Schlosser M., Kremling H., Keime S.,

RA Engel W.;

RL Eur. J. Biochem. 207:27-28(1992).

RN [5]

RP SEQUENCE OF 190-421 FROM N.A.

RA Blakey S.;

RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.

CC FUNCTION: ACROSIN IS THE MAJOR PROTEASE OF MAMMALIAN SPERMATOZOEA.

CC IT IS A SERINE PROTEASE OF TRYPSIN-LIKE CLEAVAGE SPECIFICITY, IT
 CC IS SYNTHESIZED IN A ZYMOGEN FORM, PROACROSIN AND STORED IN THE
 CC ACROSOME.

CC CATALYTIC ACTIVITY: HYDROLYSIS OF ARG- AND LYS-BONDS; PREFERENTIAL

CC CLEAVAGE ARG-XAA >> LYS-LYS >> LYS-XAA.

CC SUBUNIT: HEAVY CHAIN (CATALYTIC) AND A LIGHT CHAIN LINKED BY TWO
 CC DISULFIDE BONDS.

CC SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPSIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; Y00370; CAA68784.1; -

DR EMBL; X54017; CAA37964.1; -

DR EMBL; X54018; CAA37964.1; JOINED.

DR EMBL; X54019; CAA37964.1; JOINED.

DR EMBL; X54020; CAA37964.1; JOINED.

DR EMBL; M77378; AAA51572.1; -

DR EMBL; M77379; AAA51573.1; -

DR EMBL; M77380; AAA51574.1; -

DR EMBL; M77381; AAA51575.1; -

DR EMBL; X66188; CAA46956.1; -

DR EMBL; X54018; CAA46956.1; JOINED.

DR EMBL; X54019; CAA46956.1; JOINED.

DR EMBL; X54020; CAA46956.1; JOINED.

DR EMBL; AL078621; CAB81647.1; -

DR PIR; S03330; S03330.

DR PIR; S11674; S11674.

DR PIR; S12063; S12063.

DR PIR; S23499; S23499.

DR HSSP; P00766; 1CHG.

DR MEROPS; S01.223; -

DR MIM; 102480; -

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00089; trypsin; 1.

DR SMART; PR00722; CHYMOTRYPSIN.

DR PRINTS; SM00020; Tryp_SPC; 1.

DR PROSITE; PS0240; TRYPSIN_DOM; 1.

DR PROSITE; PS00134; TRYPSIN_HIS; 1.

DR PROSITE; PS00135; TRYPSIN_SER; 1.

KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Sperm; Signal;

KW Polymorphism.

FT SIGNAL 1 19

FT CHAIN 20 421 ACROSIN.

FT CHAIN 20 42 ACROSIN LIGHT CHAIN.

FT CHAIN 43 ? ACROSIN HEAVY CHAIN.

FT PROPEP ? 421 PRO-RICH.

FT DOMAIN 43 290 SERINE PROTEASE.

FT DISULFID 25 154 INTERCHAIN (BY SIMILARITY).

FT DISULFID 29 162 INTERCHAIN (BY SIMILARITY).

FT DISULFID 73 89 BY SIMILARITY.

FT DISULFID 177 246 BY SIMILARITY.

FT DISULFID 209 225 BY SIMILARITY.

FT DISULFID 236 266 BY SIMILARITY.

FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
 FT ACT_SITE 88 88 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 142 142 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 240 240 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT VARIANT 120 120 L -> V (IN DBSNP:1064734).
 FT VARIANT 166 166 /FTID=VAR_011650.
 FT VARIANT 166 166 F -> L (IN DBSNP:1064735).
 FT CONFLICT 64 64 /FTID=VAR_011651.
 FT CONFLICT 195 195 R -> T (IN REF. 1).
 FT CONFLICT 217 218 I -> M (IN REF. 5).
 FT CONFLICT 254 254 GR -> EH (IN REF. 5).
 FT CONFLICT 268 268 Y -> C (IN REF. 5).
 FT CONFLICT 320 320 L -> R (IN REF. 1 AND 5).
 FT CONFLICT 320 320 I -> L (IN REF. 5).
 FT CONFLICT 345 345 P -> R (IN REF. 1 AND 5).
 FT CONFLICT 358 358 A -> P (IN REF. 5).
 SQ SEQUENCE 421 AA; 45799 MW; 62EE47DC25B4FB5D CRC64;

Query Match 74.5%; Score 35; DB 1; Length 421;

Best Local Similarity 75.0%; Pred No. 27;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 LVNERWVL 9

Db 77 LLNSRWVL 84

RESULT 38

CIS_HUMAN

ID CIS_HUMAN STANDARD; PRT; 688 AA.

AC P09871; Q9UM14; Q9UCV7; Q9UCU8; Q9UCU9; Q9UCV0; Q9UCV1; Q9UCV2;

AC Q9UCV3; Q9UCV4; Q9UCV5;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-MAR-1989 (Rel. 10, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Complement C1s component precursor (EC 3.4.21.42) (C1 esterase).

GN C1S.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI_TaxID=9606;

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE=89017187; PubMed=2459702;

RA Kusumoto H., Hirose S., Saller J.P., Hagen F.S., Kurachi K.;

RT "Human genes for complement components C1r and C1s in a close

tail-to-tail arrangement.";

RL Proc. Natl. Acad. Sci. U.S.A. 85:7307-7311(1988).

[2]

RP SEQUENCE FROM N.A.

RC TISSUE=Liver;

RX MEDLINE=88082788; PubMed=3500856;

RA McKinnon C.M., Carter P.E., Smyth S.J., Dunbar B., Fothergill J.E.;

RT "Molecular cloning of cDNA for human complement component C1s. The

complete amino acid sequence.";

RL Eur. J. Biochem. 169:547-553(1987).

[3]

RP SEQUENCE FROM N.A.

RX MEDLINE=88163522; PubMed=2831944;

RA Tosi M., Duponchel C., Meo T., Julier C.;

RT "Complete cDNA sequence of human complement C1s and close physical

linkage of the homologous genes C1r and C1s.";

RL Biochemistry 26:8516-8524(1987).

[4]

RP SEQUENCE OF 1-329 FROM N.A.

RC TISSUE=Peripheral blood leukocytes;

RX MEDLINE=99008558; PubMed=9794427;

RA Endo Y., Takahashi M., Nakao M., Saiga H., Sekine H., Matsushita M.;

RA Nonaka M., Fujita T.;

RT "Two lineages of mannose-binding lectin-associated serine protease

(MASP) in vertebrates.";

RL J. Immunol. 161:4924-4930(1998).

101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

```
Db 464 LINEYVWL 471
|:|:| |
RESULT 39
YQAC_BACSU
ID YQAC_BACSU STANDARD; PRT; 178 AA.
AC P45900;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 20.7 kDa protein in SPOIIC-CWLA intergenic region
DE precursor.
GN YQAC.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=95219086; PubMed=7704261;
RA Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.;
RT "Complete nucleotide sequence of a skin element excised by DNA
rearrangement during sporulation in Bacillus subtilis.";
RL Microbiology 141:323-327(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S., Sato T.,
RA Takeuchi M.;
RL Submitted (May-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION
RX MEDLINE=96084975; PubMed=7489895;
RA Medigue C., Moszer I., Viari A., Danchin A.;
RT "Analysis of a Bacillus subtilis genome fragment using a co-operative
computer system prototype.";
RL Gene 165:GC37-GC51(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D32216; BAA06916.1; -
DR EMBL; D84432; BAA12377.1; -
DR EMBL; Z99117; CAB14578.1; -
DR Subtilist; BG11254; yqac.
KW Hypothetical protein; Signal; Complete proteome.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 178 HYPOTHETICAL PROTEIN YQAC.
SQ SEQUENCE 178 AA; 20702 MW; DD2DE09D65CF882E CRC64;
Query Match 72.3%; Score 34; DB 1; Length 178;
Best Local Similarity 57.1%; Pred. No. 17;
Matches 4; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 3 VNERVWL 9
|:|:| |
Db 61 INEKWL 67
RESULT 40
VSPA_LACMU
ID VSPA_LACMU STANDARD; PRT; 228 AA.
AC P33589;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Gyroxin analog (EC 3.4.21.74) (Thrombin-like enzyme) (Venombin A).
OS Lachesis muta muta (Bushmaster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidae;
OC Viperidae; Crotalinae; Lachesis.
OX NCBI_TaxID=8753;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=93359030; PubMed=8354384;
RA Magalhaes A., Campos Brasil da Fonseca B., Ribeiro Diniz C.,
RA Gilroy J., Richardson M.;
RT "The complete amino acid sequence of a thrombin-like enzyme/gyroxin
analogue from venom of the bushmaster snake (Lachesis muta muta).";
RL FEBS Lett. 329:116-120(1993).
RN [2]
RP SEQUENCE OF 1-25.
RC TISSUE=Venom;
RX MEDLINE=89388830; PubMed=2781576;
RA da Silva N.J., Alrd S.D., Seebart C., Kaiser I.I.;
RT "A gyroxin analog from the venom of the bushmaster (Lachesis muta
muta).";
RL Toxicon 27:763-771(1989).
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. DISPLAYS A
CC SPECIFICITY SIMILAR TO TRYPSIN. RELEASES ONLY FIBRINOPEPTIDE A IN
CC THE CONVERSION OF FIBRINOGEN TO FIBRIN.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-GLY BONDS IN
CC FIBRINOGEN ALPHA CHAINS (DOES NOT CLEAVE FIBRINOGEN BETA CHAINS).
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- MISCELLANEOUS: INHIBITED COMPETITIVELY BY AMIDINES AND GUANIDINES,
CC AND IRREVERSIBLY INHIBITED BY DIISOPROPYLFLUOROPHOSPHATE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
DR PIR; A32415; A32415.
DR PIR; S35689; S35689.
DR HSSP; P00763; LDPO.
DR MEROPS; S01.181; -.
DR InterPro; IP8001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; FALSE_NEG.
DR PROSITE; PS00135; TRYPSIN_SER; FALSE_NEG.
KW Hydrolase; Serine protease; Glycoprotein; Venom.
FT ACT_SITE 43 43
FT ACT_SITE 88 88 CHARGE RELAY SYSTEM.
FT ACT_SITE 177 177 CHARGE RELAY SYSTEM.
FT DISULFID 7 138 BY SIMILARITY.
FT DISULFID 28 44 BY SIMILARITY.
FT DISULFID 78 227 BY SIMILARITY.
FT DISULFID 117 183 BY SIMILARITY.
FT DISULFID 149 162 BY SIMILARITY.
FT DISULFID 173 198 BY SIMILARITY.
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .).
FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .).
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .).
FT CARBOHYD 224 224 N-LINKED (GLCNAC. . .).
SQ SEQUENCE 228 AA; 25629 MW; DA52305D2995C52D CRC64;
Query Match 72.3%; Score 34; DB 1; Length 228;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 2 LVNERVWL 9
|:|:| |
Db 32 LINOEWVL 39
RESULT 41
VSP1_AKCO
ID VSP1_AKCO STANDARD; PRT; 231 AA.
AC P09872;
```

DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Anctrod (EC 3.4.21.74) (Venomom A) (Protein C activator) (ACC-C).
 OS Agkistrodon contortrix contortrix (Southern copperhead).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Serpentes; Colubroidea;
 OC Viperidae; Crotalinae; Agkistrodon.
 NCBI_TaxID=8713;
 RN [1]
 RP SEQUENCE.
 RC TISSUE-Venom;
 RX MEDLINE=89229065; PubMed=2653426;
 RA McMullen B.A., Fujikawa K., Kiesel W.;
 RT "Primary structure of a protein C activator from Agkistrodon
 contortrix contortrix venom.";
 RL Biochemistry 28:674-679(1989).
 [2]
 RP SEQUENCE OF 1-63.
 RC TISSUE-Venom;
 RX MEDLINE=87308291; PubMed=3624272;
 RA Kiesel W., Kondo S., Smith K.J., McMullen B.A., Smith L.F.;
 RT "Characterization of a protein C activator from Agkistrodon
 contortrix contortrix venom";
 RL J. Biol. Chem. 262:12607-12613(1987).
 CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. CLEAVES
 FIBRINOPEPTIDES AM, AO, AND AV; THE ABERRANT FIBRINOGEN IS THEN
 INCAPABLE OF BEING CROSS-LINKED, FORMING EASILY DISPERSIBLE
 CLOTS. ACTIVATES PROTEIN C.
 CC -!- CATALYTIC ACTIVITY: Selective cleavage of Arg-I-Xaa bond in
 fibrinogen, to form fibrin, and release fibrinopeptide A. The
 specificity of further degradation of fibrinogen varies with
 species origin of the enzyme.
 CC -!- SUBCELLULAR LOCATION: Secreted (Potential).
 CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
 DR PIR: A27370; A27370.
 DR PIR: A60468; A60468.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.178; -.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: SM00020; Tryp_SPC; 1.
 DR SMART: PS0240; TRYPsin_DOM; 1.
 DR PROSITE: PS0134; TRYPsin_HIS; 1.
 DR PROSITE: PS0135; TRYPsin_SER; 1.
 KW Hydrolase; Serine protease; Venom; Glycoprotein.
 FT ACT_SITE 40 40 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 85 85 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 177 177 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 7 138 BY SIMILARITY.
 FT DISULFID 25 41 BY SIMILARITY.
 FT DISULFID 73 229 BY SIMILARITY.
 FT DISULFID 117 183 BY SIMILARITY.
 FT DISULFID 149 162 BY SIMILARITY.
 FT DISULFID 173 198 BY SIMILARITY.
 FT CARBOHYD 21 21 N-LINKED (GLCNAC. . .) (PROBABLE).
 FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .).
 SQ SEQUENCE 231 AA; 25106 MW; DC3D7CBE601EC52B CRC64;

Query Match 72.3%; Score 34; DB 1; Length 231;
 Best Local Similarity 62.5%; Pred. No. 22;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 LVNERWVL 9
 |::|::|::|
 Db 29 LINGEWVL 36

RESULT 42
 VSP1_BOTJA STANDARD; PRT; 232 AA.
 AC P81624;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Platelet-aggregating proteinase PA-BJ (EC 3.4.21.-).
 OS Bothrops jararaca (Jararaca).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Serpentes; Colubroidea;
 OC Viperidae; Crotalinae; Bothrops.
 NCBI_TaxID=8724;
 RN [1]
 RP SEQUENCE, AND CHARACTERIZATION.
 RC TISSUE-Venom;
 RX MEDLINE=95284054; PubMed=7766629;
 RA Serrano S.M.T., Mentele R., Sampaio C.A.M., Fink E.;
 RT "Purification, characterization, and amino acid sequence of a serine
 proteinase, PA-BJ, with platelet-aggregating activity from the venom
 of Bothrops jararaca.";
 RL Biochemistry 34:7186-7193(1995).
 CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. POSSESSES
 AMIDOLYTIC ACTIVITIES. CAUSES THE AGGREGATION OF PLATELETS IN
 PLATELET RICH PLASMA AND IN SUSPENSIONS OF WASHED PLATELETS.
 CC DOES NOT HAVE FIBRINOGEN-CLOTTING ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-
 CC -!- ENZYME REGULATION: INHIBITED BY PHENYLTHANESULFONYL FLUORIDE.
 CC THE AMIDOLYTIC ACTIVITY IS ALSO INHIBITED BY BENZAMIDINE
 DERIVATES.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
 DR HSSP: P00763; IDPO.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; Tryp_SPC; 1.
 DR PROSITE: PS50240; TRYPsin_DOM; 1.
 DR PROSITE: PS0134; TRYPsin_HIS; 1.
 DR PROSITE: PS0135; TRYPsin_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Venom.
 FT ACT_SITE 41 41 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 86 86 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 180 180 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 7 139 BY SIMILARITY.
 FT DISULFID 26 42 BY SIMILARITY.
 FT DISULFID 74 231 BY SIMILARITY.
 FT DISULFID 118 186 BY SIMILARITY.
 FT DISULFID 150 165 BY SIMILARITY.
 FT DISULFID 176 200 BY SIMILARITY.
 FT CARBOHYD 20 20 N-LINKED (GLCNAC. . .).
 FT CARBOHYD 23 23 O-LINKED.
 SQ SEQUENCE 232 AA; 25215 MW; 2AFE32865C4B1E74 CRC64;

Query Match 72.3%; Score 34; DB 1; Length 232;
 Best Local Similarity 62.5%; Pred. No. 22;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 2 LVNERWVL 9
 |::|::|::|
 Db 30 LINGEWVL 37

RESULT 43
 VSPA_BOTJA STANDARD; PRT; 232 AA.
 AC P81661;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Bothrombin (EC 3.4.21.74) (Reptilase).
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
RP SEQUENCE.
RC TISSUE=Venom;
RX MEDLINE=94153945; PubMed=8110787;
RA Nishida S., Fujimura Y., Miura S., Ozaki Y., Usami Y., Suzuki M.,
RA Titani K., Yoshida E., Sugimoto M., Yoshioka A., Fukui H.;
RT "Purification and characterization of bothrombin, a fibrinogen-
RT clotting serine protease from the venom of Bothrops jararaca.";
RL Biochemistry 33:1843-1849(1994).
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. AGGREGATES
CC PLATELETS IN THE PRESENCE OF EXOGENOUS FIBRINOGEN. POSSIBLY
CC THROUGH INTERACTION WITH GLYCOPROTEIN 1B. ACTIVATES BLOOD
CC COAGULATING FACTOR VIII.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-GLY BONDS IN
CC FIBRINOGEN ALPHA CHAINS (DOES NOT CLEAVE FIBRINOGEN BETA CHAINS).
CC -!- ENZYME REGULATION: INHIBITED BY DIISOPROPYL FLUOROPHOSPHATE, BUT
CC NOT BY HIRUDIN.
CC -!- SUBUNIT: MONOMER.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Hydrolase: Serine protease; Glycoprotein; Venom.
KW ACT_SITE 41 41 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 86 86 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 178 178 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 7 139 BY SIMILARITY.
FT DISULFID 26 42 BY SIMILARITY.
FT DISULFID 74 230 BY SIMILARITY.
FT DISULFID 118 184 BY SIMILARITY.
FT DISULFID 150 163 BY SIMILARITY.
FT DISULFID 174 199 BY SIMILARITY.
FT CARBOHYD 98 98 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 146 146 N-LINKED (GLCNAC. . .) (PROBABLE).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (PROBABLE).
SQ SEQUENCE 232 AA; 25584 MW; BFCDA6A729D3B8AA CRC64;

Query Match 72.3%; Score 34; DB 1; Length 232;
Best Local Similarity 62.5%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I: I: I I I
DB 30 LINQEWVL 37

RESULT 44

VSP2_AGKB1
ID VSP2_AGKB1 STANDARD; PRT; 235 AA.
AC Q9PSN3; Q9PS18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bilineobin (EC 3.4.21.-).
OS Agkistrodon bilineatus (Cantil) (Tropical moccasin).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Agkistrodon.
OX NCBI_TaxID=8718;

RN [1]
RP SEQUENCE, AND DISULFIDE BONDS.
RC TISSUE=Venom;
RX MEDLINE=95243661; PubMed=7726578;
RA Nikai T., Ohara A., Komori Y., Fox J.W., Sugihara H.;
RT "Primary structure of a coagulant enzyme, bilineobin, from Agkistrodon
RT bilineatus venom.";
RL Arch. Biochem. Biophys. 318:89-96(1995).
RN [2]
RP PARTIAL SEQUENCE OF 1-24, AND CHARACTERIZATION.
RC TISSUE=Venom;
RX MEDLINE=93227264; PubMed=8470131;
RA Komori Y., Nikai T., Ohara A., Yagihashi S., Sugihara H.;
RT "Effect of bilineobin, a thrombin-like proteinase from the venom of
RT common cantil (Agkistrodon bilineatus).";
RL Toxicon 31:257-270(1993).
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. HAS A
CC COAGULANT ACTIVITY. ACTS ON FIBRINOGEN ALPHA AND BETA GENERATING
CC FIBRINOPEPTIDES A AND B.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
DR HSSP: P00752; 2PKA.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
DR Hydrolase: Serine protease; Glycoprotein; Venom.
KW ACT_SITE 43 43 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 88 88 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 182 182 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 7 141 N-LINKED (GLCNAC. . .).
FT DISULFID 28 44 N-LINKED (GLCNAC. . .).
FT DISULFID 78 234 N-LINKED (GLCNAC. . .).
FT DISULFID 120 188 N-LINKED (GLCNAC. . .).
FT DISULFID 152 167 N-LINKED (GLCNAC. . .).
FT DISULFID 178 203 N-LINKED (GLCNAC. . .).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .).
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .).
FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .).
FT CARBOHYD 132 132 N-LINKED (GLCNAC. . .).
FT CARBOHYD 148 148 N-LINKED (GLCNAC. . .).
FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .).
FT CONFLICT 1 1 I -> V (IN REF. 2).
SQ SEQUENCE 235 AA; 26478 MW; 73A98766CA507578 CRC64;

Query Match 72.3%; Score 34; DB 1; Length 235;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I: I: I I I
DB 32 LINQEWVL 39

RESULT 45

VSP1_AGKHA
ID VSP1_AGKHA STANDARD; PRT; 238 AA.
AC P81176;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Halystase (EC 3.4.21.-).
OS Agkistrodon halys blomhoffii (Mamushi) (Gloydius blomhoffii).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Agkistrodon.
OX NCBI_TaxID=61300;

```
RN [1]
RP SEQUENCE.
RC TISSUE-Venom;
RX MEDLINE=98206752; PubMed=9546675;
RA Matsui T., Sakurai Y., Fujimura Y., Hayashi I., Oh-Ishi S., Suzuki M.,
RA Hamako J., Yamamoto Y., Yamazaki J., Kinoshita M., Titani K.;
RT "Purification and amino acid sequence of halystase from snake venom
RT of Agkistrodon halys blomhoffii, a serine protease that cleaves
RT specifically fibrinogen and kininogen.";
RL Eur. J. Biochem. 252:569-575(1998);
CC -!- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE. CLEAVES
CC FIBRINOGEN WITHOUT INDUCING FIBRIN CLOTTING AND CLEAVES
CC KININOGEN TO PRODUCE BRADYKININ, RESULTING IN THE REDUCTION
CC OF BLOOD PRESSURE.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
DR HSP; P00752; 2PKA.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Venom; Glycoprotein.
FT ACT_SITE 43 43 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 88 88 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 184 184 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 7 141 BY SIMILARITY.
FT DISULFID 28 44 BY SIMILARITY.
FT DISULFID 76 236 BY SIMILARITY.
FT DISULFID 120 190 BY SIMILARITY.
FT DISULFID 152 169 BY SIMILARITY.
FT DISULFID 180 205 BY SIMILARITY.
FT CARBOHYD 81 81 N-LINKED (GLCNAC. . .).
FT CARBOHYD 100 100 N-LINKED (GLCNAC. . .).
FT VARIANT 21 23 PRS -> SRV.
SQ SEQUENCE 238 AA; 26483 MW; 00C658067DE27586 CRC64;

Query Match 72.3%; Score 34; DB 1; Length 238;
Best Local Similarity 62.5%; Pred. No. 23;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 1:1: |||
32 LINQEWVL 39
```

Search completed: November 6, 2002, 12:09:09
Job time : 7.33333 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 11 Seconds

(without alignments)
78.619 Million cell updates/sec

Title: US-09-905-083-34

Perfect score: 47

Sequence: 1 VLVNERWVL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	100.0	253	2 A53968	serine proteinase
2	39	83.0	246	1 DBHU	complement factor
3	38	80.9	232	1 KQPG	tissue kallikrein
4	38	80.9	260	2 I56559	neuropsin - mouse
5	38	80.9	417	2 T25594	hypothetical prote
6	37	78.7	250	2 T01779	trypsin (EC 3.4.21
7	37	78.7	250	2 S31384	trypsin (EC 3.4.21
8	37	78.7	258	1 S44184	chymotrypsin (EC 3
9	37	78.7	259	1 S49129	chymotrypsin (EC 3
10	37	78.7	1421	2 T05892	hypothetical prote
11	36	76.6	72	2 F82825	hypothetical prote
12	36	76.6	258	2 A45161	serine proteinase
13	36	76.6	426	2 T28563	hypothetical prote
14	36	76.6	426	2 C36850	A21R protein - var
15	36	76.6	426	2 C72166	A22R protein - var
16	36	76.6	426	2 T37408	probable 49.1k pro
17	36	76.6	426	2 D42519	A20R protein - vac
18	36	76.6	516	2 T47075	hypothetical prote
19	35	74.5	178	2 T25570	hypothetical prote
20	35	74.5	231	2 S31778	trypsin (EC 3.4.21
21	35	74.5	242	2 S31775	trypsin (EC 3.4.21
22	35	74.5	242	2 S49489	trypsin (EC 3.4.21
23	35	74.5	242	2 S31776	trypsin (EC 3.4.21
24	35	74.5	247	1 A25852	trypsin (EC 3.4.21
25	35	74.5	257	1 JC2479	venombin B (EC 3.4
26	35	74.5	258	2 I36945	haptoglobin Hp - c
27	35	74.5	258	2 I36947	haptoglobin Hpp -
28	35	74.5	258	2 G02959	haptoglobin - rhes
29	35	74.5	274	2 I47078	coagulation factor

30 74.5 274 2 E70472
31 35 282 2 AB3757
32 35 74.5 329 1 HPGS
33 35 74.5 345 2 I36941
34 35 74.5 346 2 I36942
35 35 74.5 347 1 HPHU1
36 35 74.5 347 2 G00006
37 35 74.5 348 1 HPHUR
38 35 74.5 349 2 I36944
39 35 74.5 406 1 HPHU2
40 35 74.5 410 2 T46034
41 35 74.5 416 1 KFBO
42 35 74.5 421 1 S11674
43 35 74.5 453 2 G96695
44 35 74.5 505 2 C86443
45 35 74.5 539 2 S54045

ALIGNMENTS

RESULT 1

A53968

serine proteinase SCCE precursor - human

N:Alternate names: stratum corneum chymotryptic enzyme

C:Species: Homo sapiens (man)

C>Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999

C:Accession: A53968

R:Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud,

J. Biol. Chem. 269, 19420-19426, 1994

A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic en

A:Reference number: A53968; MUID:94308225

A:Accession: A53968

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-253 <HAN>

A:Cross-references: GB:I33404; NID:g521214; PIDN:AAC37551.1; PID:g532504

C:Genetics:

A:Gene: GDB:PRSS6; SCCE

A:Cross-references: GDB:377730

A:Map position: 7q35-7q35

C:Superfamily: trypsin; trypsin homology

F:30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 47; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.25; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0

QY 1 VLVNERWVL 9

|||||

Db 58 VLVNERWVL 66

RESULT 2

DBHU

complement factor D (EC 3.4.21.46) precursor [validated] - human (fragment)

N:Alternate names: adipsin; C3 convertase activator

C:Species: Homo sapiens (man)

C>Date: 28-Aug-1985 #sequence_revision 31-Dec-1992 #text_change 08-Dec-2000

C:Accession: A40197; A00936; A60571; S66645

R:White, R.T.; Damm, D.; Hancock, N.; Rosen, B.S.; Lowell, B.B.; Usher, P.; Flier, J

J. Biol. Chem. 267, 9210-9213, 1992

A:Title: Human adipsin is identical to complement factor D and is expressed at high

A:Reference number: A40197; MUID:92250520

A:Accession: A40197

A:Molecule type: mRNA

A:Residues: 1-246 <WHI>

A:Cross-references: GB:M84526

R:Niemann, M.A.; Bhowm, A.S.; Bennett, J.C.; Volanakis, J.E.

Biochemistry 23, 2482-2486, 1984

A:Title: Amino acid sequence of human D of the alternative complement pathway.

A:Reference number: A00936; MUID:85000441

A:Accession: A00936

A:Molecule type: protein
A:Residues: 19-44,'G',46-51,'Q',53-75,'TH',78,'P',80-83,'XXXITIE',90-172,86-91,185-235,'
A:Note: a few residues were assigned from the previously published sequence of Reid et al
R:Miyaata, T.; Oda, T.; Inagi, R.; Sugiyama, S.; Miyama, A.; Maeda, K.; Nakashima, I.; Ya
Mol. Immunol. 27, 637-644, 1990
A:Title: Molecular and functional identification and purification of complement component
A:Reference number: A60571; MUID:90370044
A:Accession: A60571
A:Molecule type: protein
A:Residues: 19-20,'XX',23-27,'XX',30-31,'XX',34,'X',36-40 <MIY>
R:Balke, N.; Holtkamp, U.; Hoerl, W.H.; Tschesche, H.
FEBS Lett. 371, 300-302, 1995
A:Title: Inhibition of degranulation of human polymorphonuclear leukocytes by complement
A:Reference number: S66645; MUID:96013156
A:Accession: S66645
A>Status: preliminary
A:Molecule type: protein
A:Residues: 19-44,'C',46-48 <BAL>
C:Comment: Factor D cleaves factor B when the latter is complexed with factor C3b, activ
C:Genetics:
A:Gene: GDB:DF
A:Cross-references: GDB:132645; OMIM:134350
A:Map position: Xpter-Xqter
C:Superfamily: trypsin; trypsin homology
C:Keywords: complement alternate pathway; hydrolase; plasma; serine proteinase
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-246/Product: complement factor D (fragment) #status experimental <MAT>
F:19-241/Domain: trypsin homology <TRY>
F:44-60,141-207,172-188,197-222/Disulfide bonds: #status predicted
F:59,105,201/Active site: His, Asp, Ser #status predicted

Query Match 83.0%; Score 39; DB 1; Length 246;
Best Local Similarity 88.9%; Pred. No. 7.5;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||| |||||
Db 47 VLVAERWVL 55

RESULT 3
KQPG
tissue kallikrein (EC 3.4.21.35), pancreatic - pig (tentative sequence)
N:Alternate names: glandular kallikrein; kininogenin
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 24-Apr-1984 #sequence_revision 31-Dec-1993 #text_change 31-Mar-2000
A:Accession: A00938; A92895
R:Tschesche, H.; Mair, G.; Godec, G.; Fiedler, F.; Ehret, W.; Hirschauer, C.; Lemon, M.;
Adv. Exp. Med. Biol. 120, 245-260, 1979
A:Title: The primary structure of porcine glandular kallikreins.
A:Reference number: A90015
A:Accession: A00938
A:Molecule type: protein
A:Residues: 1-49,'GWL',53-134,'D',136-156,'H',158,'B',160-224,'B',226-232 <TSC>
A:Note: The residue identified as 225-Asx is bound to carbohydrate; therefore, we have s
R:Bode, W.; Chen, Z.; Bartels, K.; Kutzbach, C.; Schmidt-Kastner, G.; Bartunik, H.
J. Mol. Biol. 164, 237-282, 1983
A:Title: Refined 2 angstrom X-ray crystal structure of porcine pancreatic kallikrein A,
t, structure and its comparison with bovine trypsin.
A:Reference number: A92895; MUID:83189107
A:Contents: X-ray crystallography, 2 angstroms
A:Accession: A92895
A:Molecule type: protein
A:Residues: 1-224,'B',226-232 <BOD>
C:Comment: The protein consists of two chains, A and B, held together by disulfide bonds
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; pancreas; serine proteinase
F:1-224/Domain: trypsin homology <TRY>
F:1-80/Product: tissue kallikrein chain A #status experimental <MPTA>
F:79-82/Region: autolysis loop
F:81-232/Product: tissue kallikrein chain B #status experimental <MPTB>
F:7-144,26-42,121-190,155-169,180-205/Disulfide bonds: #status experimental

F:41,89,184/Active site: His, Asp, Ser #status experimental
F:78/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:225/Binding site: carbohydrate (Asn) (covalent) (partial) #status experimental

Query Match 80.9%; Score 38; DB 1; Length 232;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||| |||||
Db 29 VLVNPKRWVL 37

RESULT 4
I56559
neuropsin - mouse
C:Species: Mus musculus (house mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C:Accession: I56559
R:Chen, Z.L.; Yoshida, S.; Kato, K.; Momota, Y.; Suzuki, J.; Tanaka, T.; Ito, J.; Nis
J. Neurosci. 15, 5088-5097, 1995
A:Title: Expression and activity-dependent changes of a novel limbic-serine protease
A:Reference number: I56559; MUID:95348617
A:Accession: I56559
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-260 <RES>
A:Cross-references: GB:D30785; NID:g1648847; PIDN:BAA06451.1; PID:g1020091
C:Superfamily: trypsin; trypsin homology
F:33-252/Domain: trypsin homology <TRY>

Query Match 80.9%; Score 38; DB 2; Length 260;
Best Local Similarity 77.8%; Pred. No. 12;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||| |||||
Db 61 VLVGDRWVL 69

RESULT 5
T25594
hypothetical protein C32E8.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25594
R:Gatting, S.
submitted to the EMBL Data Library, February 1997
A:Description: The sequence of C. elegans cosmid C32E8.
A:Reference number: Z20056
A:Accession: T25594
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-417 <GAP>
A:Cross-references: EMBL:U88308; PIDN:AAB42324.1; GSPDB:GN000019; CESP:C32E8.6
A:Experimental source: strain Bristol N2; clone C32E8
C:Genetics:
A:Gene: CESP:C32E8.6
A:Map position: 1
A:Introns: 22/3; 51/2; 92/1; 297/1; 354/1

Query Match 80.9%; Score 38; DB 2; Length 417;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERW 7
||| |||||
Db 219 VLINERW 225

RESULT 6
T01779
trypsin (EC 3.4.21.4) - plaice

C:Species: Pleuronectes platessa (plaice)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 16-Jul-1999

C:Accession: T01779
R:Leaver, M.J.; George, S.G.
submitted to the EMBL Data Library, March 1996
A:Reference number: Z14422
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-250 <LEA>
A:Cross-references: EMBL:X56744; NID:g1213630; PID:g64240
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; protein digestion; serine proteinase
F:23-242/Domain: trypsin homology <TRY>

Query Match 78.7%; Score 37; DB 2; Length 250;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||:|:|
Db 50 VLINQWVL 58

RESULT 7

S31384

trypsin (EC 3.4.21.4) - plaice

C:Species: Pleuronectes platessa (plaice)
C:Date: 02-Dec-1993 #sequence_revision 03-Aug-1995 #text_change 18-Jul-1997
C:Accession: S31384

R:Leaver, M.J.; George, S.G.
submitted to the EMBL Data Library, November 1990
A:Reference number: S30567

A:Accession: S31384

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-250 <LEA>

A:Cross-references: EMBL:X56744

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; protein digestion; serine proteinase

F:23-242/Domain: trypsin homology <TRY>

Query Match 78.7%; Score 37; DB 2; Length 250;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||:|:|
Db 50 VLINQWVL 58

RESULT 8

S4184

chymotrypsin (EC 3.4.21.1) 2 precursor - African malaria mosquito

N:Alternate names: chymotrypsin-like proteinase ANCHYM2

C:Species: Anopheles gambiae (African malaria mosquito)

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
C:Accession: S4184; S49130

R:MueUeller, H.M.; Catteruccia, F.; Crisanti, A.

submitted to the EMBL Data Library, April 1994

A:Description: An Anopheles gambiae locus containing the sequences of two closely related

A:Reference number: S44184

A:Accession: S44184

A:Molecule type: DNA

A:Residues: 1-258 <MUE>

A:Cross-references: EMBL:Z32645; NID:g474026; PIDN:CAA83567.1; PID:g474027

R:MueUeller, H.

submitted to the EMBL Data Library, November 1992

A:Description: Cloning of blood meal induced serine protease genes of the mosquito Anoph

A:Reference number: S35412

A:Accession: S49130

A:Molecule type: DNA

A:Residues: 1-165,'R',167-193,'PPD',197-258 <MU2>

A:Cross-references: EMBL:Z18888; NID:g509416; PIDN:CAA79326.1; PID:g509417
C:Genetics:

A:Introns: 78/2; 209/3

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; protein digestion; serine proteinase

F:1-17/Domain: signal sequence #status predicted <SIG>

F:18-32/Domain: activation peptide #status predicted <PRO>

F:33-258/Product: chymotrypsin 2 #status predicted <MAT>

F:33-250/Domain: trypsin homology <TRY>

F:59-75,182-198,208-232/Disulfide bonds: #status experimental

F:74,119,212/Active site: His, Asp, Ser #status predicted

Query Match 78.7%; Score 37; DB 1; Length 258;

Best Local Similarity 75.0%; Pred. No. 19;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

||:|:|

Db 63 LLNDRWVL 70

RESULT 9

S49129

chymotrypsin (EC 3.4.21.1) 1 precursor - African malaria mosquito

N:Alternate names: chymotrypsin-like proteinase ANCHYM1

C:Species: Anopheles gambiae (African malaria mosquito)

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000

C:Accession: S49129; S44185

R:MueUeller, H.

submitted to the EMBL Data Library, November 1992

A:Description: Cloning of blood meal induced serine protease genes of the mosquito An

A:Reference number: S35412

A:Accession: S49129

A:Molecule type: mRNA

A:Residues: 1-259 <MUE>

A:Cross-references: EMBL:Z18887; NID:g509414; PIDN:CAA79325.1; PID:g509415

A:Experimental source: strain Suakoko

R:MueUeller, H.M.; Catteruccia, F.; Crisanti, A.

submitted to the EMBL Data Library, April 1994

A:Description: An Anopheles gambiae locus containing the sequences of two closely rel

A:Reference number: S44184

A:Accession: S44185

A:Molecule type: DNA

A:Residues: 1-259 <MU2>

A:Cross-references: EMBL:Z32645; NID:g474026; PIDN:CAA83568.1; PID:g474028

A:Experimental source: strain Suakoko

C:Genetics:

A:Introns: 78/2; 209/3

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; protein digestion; serine proteinase

F:1-17/Domain: signal sequence #status predicted <SIG>

F:18-32/Domain: activation peptide #status predicted <PRO>

F:33-259/Product: chymotrypsin 1 #status predicted <MAT>

F:33-250/Domain: trypsin homology <TRY>

F:59-75,182-198,208-232/Disulfide bonds: #status experimental

F:74,119,212/Active site: His, Asp, Ser #status predicted

Query Match 78.7%; Score 37; DB 1; Length 259;

Best Local Similarity 75.0%; Pred. No. 19;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

||:|:|

Db 63 LLNDRWVL 70

RESULT 10

T05892

hypothetical protein F6H11.110 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 13-Aug-1999

C:Accession: T05892

R:Bevan, M.; Brandt, P.; Dose, S.; Jarke, D.; Scharfe, M.; Schon, O.; Bancroft, I.; M

submitted to the Protein Sequence Database, April 1998

A:Reference number: Z15456
A:Accession: T05892
A:Molecule type: DNA
A:Residues: 1-1421 <BEV>
A:Cross-references: EMBL:AL021684
A:Experimental source: cultivar Columbia; BAC clone F6H11
C:Genetics:
A:Map position: 5
A:Introns: 72/3; 142/3; 227/2; 344/3; 372/3; 941/2; 1016/3; 1038/3; 1055/1; 1095/2; 1155
A:Note: F6H11.110

Query Match 78.7%; Score 37; DB 2; Length 1421;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||| |||
Db 1208 VLANKRWVI 1216

RESULT 11

F82825
hypothetical protein XF0279 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C:Accession: F82825
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: F82825
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-72 <SIM>
A:Cross-references: GB:AE003881; GB:AE003849; NID:g9105093; PIDN:AAF83092.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorfy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, H.M.B.N.; Madeira, A.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak
M.; Tshahko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0279

Query Match 76.6%; Score 36; DB 2; Length 72;
Best Local Similarity 62.5%; Pred. No. 7.3;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWV 8
:|:|:|:|
Db 50 ILNERNWL 57

RESULT 12

A45161
serine proteinase (EC 3.4.21.-) RNK-Met-1 - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 22-Jun-1999
C:Accession: A45161
R:Smyth, M.J.; Wiltrout, T.; Trapani, J.A.; Ottaway, K.S.; Sowder, R.; Henderson, L.E.;
J. Biol. Chem. 267, 24418-24425, 1992
A:Title: Purification and cloning of a novel serine proteinase, RNK-Met-1, from the granu

A:Reference number: A45161; MUID:93077530
A:Accession: A45161
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-258 <SMI>
A:Cross-references: GB:L05175; NID:g206691; PIDN:AAA42056.1; PID:g206692
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:21-245/Domain: trypsin homology <TRY>

Query Match 76.6%; Score 36; DB 2; Length 258;
Best Local Similarity 66.7%; Pred. No. 28;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
|||:|:|:|
Db 49 VLVHQRWVL 57

RESULT 13

T28563
hypothetical protein A21R - variola major virus
C:Species: variola major virus
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jul-2000
C:Accession: T28563
R:Massung, R.F.; Esposito, J.J.; Liu, L.I.; Qi, J.; Utterback, T.R.; Knight, J.C.; Au
Nature 366, 748-751, 1993
A:Title: Potential virulence determinants in terminal regions of variola smallpox vir
A:Reference number: Z20488; MUID:94088747
A:Accession: T28563
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-426 <MAS>
A:Cross-references: EMBL:L22579; NID:g623595; PIDN:AAA60873.1; PID:g439043
A:Experimental source: strain Bangladesh-1975
C:Superfamily: vaccinia virus probable 49.1K protein

Query Match 76.6%; Score 36; DB 2; Length 426;
Best Local Similarity 55.6%; Pred. No. 48;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||| | |:
Db 130 VLLNNRWIM 138

RESULT 14

C36850
A21R protein - variola virus (strain India-1967)
C:Species: variola virus
C:Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 23-Mar-2001
C:Accession: C36850
R:Blinov, V.M.
submitted to GenBank, November 1992

A:Reference number: A36859
A:Accession: C36850
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-426 <BLI>
A:Cross-references: GB:X69198; NID:g456758; PIDN:CAA49065.1; PID:g297303
C:Superfamily: vaccinia virus probable 49.1K protein

Query Match 76.6%; Score 36; DB 2; Length 426;
Best Local Similarity 55.6%; Pred. No. 48;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||| | |:
Db 130 VLLNNRWIM 138

RESULT 15

C72166

A22R protein - variola minor virus (strain Garcia-1966)
 C:Species: variola minor virus
 C>Date: 24-Nov-1999 #sequence_revision 24-Nov-1999 #text_change 20-Jun-2000
 C:Accession: C72166
 R:Shchelkunov, S.N.; Totmenin, A.V.; Guttorov, V.V.; Safronov, P.F.; Massung, R.F.; Lopar
 submitted to GenBank, March 1998
 A:Description: Analysis of the complete coding sequence of DNA of alastrim variola minor
 A:Reference number: A72150
 A:Accession: C72166
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-426 <SHC>
 A:Cross-references: GB:Y16780; NID:g5830555; PIDN:CAB54724.1; PID:g5830685
 A:Experimental source: strain Garcia-1966
 C:Genetics:
 A:Gene: A22R
 C:Superfamily: vaccinia virus probable 49.1K protein

Query Match 76.6%; Score 36; DB 2; Length 426;
 Best Local Similarity 55.6%; Pred. No. 48;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 II: I III:
 Db 130 VLLNNRWIM 138

RESULT 16
 T37408
 probable 49.1K protein - vaccinia virus (strain Ankara)
 C:Species: vaccinia virus
 A:Variety: strain Ankara
 C>Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jun-2000
 C:Accession: T37408
 R:Antoine, G.; Scheiflinger, F.; Falkner, F.G.; Dorner, F.
 submitted to the EMBL Data Library, March 1997
 A:Description: The complete genomic sequence of the Modified Vaccinia Ankara (MVA) strain
 A:Reference number: Z20877
 A:Accession: T37408
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-426 <ANT>
 A:Cross-references: EMBL:U94848; PIDN:AAB96523.1
 A:Experimental source: strain Ankara
 C:Genetics:
 A:Note: MVA132R
 C:Superfamily: vaccinia virus probable 49.1K protein

Query Match 76.6%; Score 36; DB 2; Length 426;
 Best Local Similarity 55.6%; Pred. No. 48;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 II: I III:
 Db 130 VLLNNRWIM 138

RESULT 17
 D42519
 A20R protein - vaccinia virus (strain Copenhagen)
 C:Species: vaccinia virus
 A:Note: host Homo sapiens (man)
 C>Date: 09-Nov-1990 #sequence_revision 09-Nov-1990 #text_change 20-Jun-2000
 C:Accession: D42519
 R:Johnson, G.P.
 submitted to GenBank, June 1990
 A:Reference number: A33172
 A:Accession: D42519
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-426 <JOH>
 C:Superfamily: vaccinia virus probable 49.1K protein

Query Match 76.6%; Score 36; DB 2; Length 426;
 Best Local Similarity 55.6%; Pred. No. 48;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 II: I III:
 Db 130 VLLNNRWIM 138

RESULT 18
 T47075
 hypothetical protein Y4kJ [imported] - Rhizobium sp. (NGR234) plasmid pNGR234a
 C:Species: Rhizobium sp.
 A:Variety: NGR234
 C>Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 17-Mar-2000
 R:Freiberg, C.; Fellay, R.; Bairoch, A.; Broughton, W.J.; Rosenthal, A.; Perret, X.
 Nature 387, 394-401, 1997
 A:Title: Molecular basis of symbiosis between Rhizobium and legumes.
 A:Reference number: Z14734; MUID:97305956
 A:Accession: T47075
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-516 <FRE>
 A:Cross-references: EMBL:AE000081; PIDN:AAB91741.1
 C:Genetics:
 A:Genome: plasmid pNGR234a
 A:Note: y4kJ

Query Match 76.6%; Score 36; DB 2; Length 516;
 Best Local Similarity 66.7%; Pred. No. 59;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 II: I III:
 Db 266 VLVVERWIL 274

RESULT 19
 T25570
 hypothetical protein C24G7.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T25570
 R:Goela, D.; Wilson, R.
 submitted to the EMBL Data Library, February 1997
 A:Description: The sequence of C. elegans cosmid C24G7.
 A:Reference number: Z20052
 A:Accession: T25570
 A>Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-178 <GOE>
 A:Cross-references: EMBL:U08310; PIDN:AAB42339.1; GSPDB:GN00019; CESP:C24G7.3
 A:Experimental source: strain Bristol N2; clone C24G7
 C:Genetics:
 A:Gene: CESP:C24G7.3
 A:Map position: 1
 A:Introns: 11/1; 65/2

Query Match 74.5%; Score 35; DB 2; Length 178;
 Best Local Similarity 55.6%; Pred. No. 29;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 II: I III:
 Db 20 IIEKWKWL 28

RESULT 20
 S31778
 trypsin (EC 3.4.21.4) II precursor - Atlantic salmon (fragment)
 C:Species: Salmo salar (Atlantic salmon)
 C>Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999

C:Accession: S66658; S31778
R:Male, R.; Lorens, J.B.; Smalas, A.O.; Torrisen, K.R.
Eur. J. Biochem. 232, 677-685, 1995
A:Title: Molecular cloning and characterization of anionic and cationic variants of trypsin
A:Reference number: S66657; MUID:96035908
A:Accession: S66658
A:Molecule type: mRNA
A:Residues: 1-231 <MAL>
A:Cross-references: EMBL:X70073; NID:g64385; PIDN:CAA49678.1; PID:g64386
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-4/Domain: signal sequence.(fragment) #status predicted <SIG>
F:5-9/Domain: activation peptide #status predicted <APT>
F:10-231/Product: trypsin II #status predicted <MAT>
F:10-224/Domain: trypsin homology <TRY>
F:16-145,34-50,118-218,125-191,156-170,181-205/Disulfide bonds: #status predicted
F:49,93,185/Active site: His, Asp, Ser #status predicted

Query Match 74.5%; Score 35; DB 2; Length 231;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
Db 38 LVNENWV 45
|||||

RESULT 21
S31775
trypsin (EC 3.4.21.4) I precursor - Atlantic salmon
C:Species: Salmo salar (Atlantic salmon)
C:Date: 03-Mar-1994 #sequence_revision 03-Aug-1995 #text_change 15-Oct-1999
C:Accession: S66660; S66661; S31775; S31777
R:Male, R.; Lorens, J.B.; Smalas, A.O.; Torrisen, K.R.
Eur. J. Biochem. 232, 677-685, 1995
A:Title: Molecular cloning and characterization of anionic and cationic variants of trypsin
A:Reference number: S66657; MUID:96035908
A:Accession: S66660
A:Molecule type: mRNA
A:Residues: 1-242 <MAL>
A:Cross-references: EMBL:X70075; NID:g64379; PIDN:CAA49680.1; PID:g64380
A:Experimental source: pancreas
A:Accession: S66661
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 28-32,'A',34-242 <MA2>
A:Cross-references: EMBL:X70072; NID:g64383; PIDN:CAA49677.1; PID:g64384
A:Experimental source: pancreas
A:Note: trypsin IB, probably an allelic variant
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-20/Domain: activation peptide #status predicted <APT>
F:21-242/Product: trypsin I #status predicted <MAT>
F:21-235/Domain: trypsin homology <TRY>
F:27-156,45-61,129-229,136-202,167-181,192-216/Disulfide bonds: #status predicted
F:60,104,196/Active site: His, Asp, Ser #status predicted

Query Match 74.5%; Score 35; DB 2; Length 242;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
Db 49 LVNENWV 56
|||||

RESULT 22
S49489
trypsin (EC 3.4.21.4) precursor - Paranotothenia magellanica
C:Species: Paranotothenia magellanica
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 22-Jun-1999
C:Accession: S49489

R:Genicot, S.; Rentier-Delrue, F.; Edwards, D.; van Beeumen, J.; Dodson, G.; Gerday, submitted to the EMBL Data Library, October 1994
A:Description: Trypsin and trypsinogen from an antarctic fish: molecular basis of col
A:Reference number: S49489
A:Accession: S49489
A:Molecule type: mRNA
A:Residues: 1-242 <GEN>
A:Cross-references: EMBL:X82223; NID:g559507; PIDN:CAA57701.1; PID:g559508
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-13/Domain: signal sequence #status predicted <SIG>
F:14-242/Product: trypsin #status predicted <MAT>
F:21-235/Domain: trypsin homology <TRY>

Query Match 74.5%; Score 35; DB 2; Length 242;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
Db 49 LVNENWV 56
|||||

RESULT 23
S31776
trypsin (EC 3.4.21.4) IA precursor - Atlantic salmon
C:Species: Salmo salar (Atlantic salmon)
C:Date: 03-Mar-1994 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C:Accession: S66659; S31776
R:Male, R.; Lorens, J.B.; Smalas, A.O.; Torrisen, K.R.
Eur. J. Biochem. 232, 677-685, 1995
A:Title: Molecular cloning and characterization of anionic and cationic variants of t
A:Reference number: S66657; MUID:96035908
A:Accession: S66659
A:Molecule type: mRNA
A:Residues: 1-242 <MAL>
A:Cross-references: EMBL:X70071; NID:g64381; PIDN:CAA49676.1; PID:g64382
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-15/Domain: signal sequence #status predicted <SIG>
F:16-20/Domain: activation peptide #status predicted <APT>
F:21-242/Product: trypsin IA #status predicted <MAT>
F:21-235/Domain: trypsin homology <TRY>
F:27-156,45-61,129-229,136-202,167-181,192-216/Disulfide bonds: #status predicted
F:60,104,196/Active site: His, Asp, Ser #status predicted

Query Match 74.5%; Score 35; DB 2; Length 242;
Best Local Similarity 75.0%; Pred. No. 41;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LVNERWVL 9
Db 49 LVNENWV 56
|||||

RESULT 24
A25852
trypsin (EC 3.4.21.4) I precursor [validated] - human
N:Alternate names: trypsin, cationic; trypsinogen I
C:Species: Homo sapiens (man)
C:Date: 03-Mar-1994 #sequence_revision 03-Mar-1994 #text_change 08-Dec-2000
C:Accession: A25852; B61066; A43988
R:Emi, M.; Nakamura, Y.; Ogawa, M.; Yamamoto, T.; Nishide, T.; Mori, T.; Matsubara, K
Gene 41, 305-310, 1986
A:Title: Cloning, characterization and nucleotide sequences of two cDNAs encoding hum-
A:Reference number: A91544; MUID:86221712
A:Accession: A25852
A:Molecule type: mRNA
A:Residues: 1-247 <EMI>
A:Cross-references: GB:M22612; NID:g521215; PIDN:AAA61231.1; PID:g521216
R:Kjmland, M.; Russick, C.; Marks, W.H.; Borgstroem, A.
Clin. Chim. Acta 184, 31-46, 1989
A:Title: Immunoreactive anionic and cationic trypsin in human serum.

A:Reference number: A61066; MUID:90091010

A:Accession: B61066

A:Molecule type: protein

A:Residues: 16-43 <KIM>

R:Koivunen, E.; Huhtala, M.L.; Stenman, U.H.

J. Biol. Chem. 264, 14095-14099, 1989

A:Title: Human ovarian tumor-associated trypsin. Its purification and characterization

A:Reference number: A43988; MUID:89340515

A:Accession: A43988

A:Molecule type: protein

A:Residues: 16-34 <KOI>

A:Experimental source: mucinous ovarian tumor cyst fluid

C:Genetics:

A:Gene: GDB:PRSS1; TRY1

A:Cross-references: GDB:119620; OMIM:276000

A:Map position: 7q35-7q35

A:Note: The human genome contains at least ten trypsin genes or pseudogenes, at least two

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; pancreas; phosphoprotein; protein digestion; serine proteinase; z

F:1-15/Domain: signal sequence #status predicted <SIG>

F:16-246/Product: trypsinogen I #status experimental <ZYM>

F:16-23/Domain: activation peptide #status experimental <APT>

F:24-246/Product: trypsin I #status predicted <ENZ>

F:24-239/Domain: trypsin homology <TRY>

F:30-160,48-64,139-206,171-185,196-220/Disulfide bonds: #status predicted

F:63,107,200/Active site: His, Asp, Ser #status predicted

F:75,77,80,85/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 74.5%; Score 35; DB 1; Length 247;

Best Local Similarity 62.5%; Pred. No. 42;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I:|:|:|:|

Db 52 LINEQWV 59

RESULT 25

JC2479

venom B (EC 3.4.21.-) precursor [validated] - Chinese habu

N:Alternate names: beta-fibrinogenase; fibrinogenolytic proteinase

C:Species: Trimeresurus mucrosquamatus (Chinese habu)

C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000

C:Accession: JC2479; A38940

R:Hung, C.C.; Huang, K.F.; Chiou, S.H.

Biochem. Biophys. Res. Commun. 205, 1707-1715, 1994

A:Title: Characterization of one novel venom protease with beta-fibrinogenase activity

A:Reference number: JC2479; MUID:95110313

A:Accession: JC2479

A:Molecule type: mRNA

A:Residues: 1-257 <HUN>

A:Cross-references: GB:X83221; NID:9602595; PIDN:CAA58221.1; PID:9602596

A:Experimental source: venom

A:Note: part of this sequence, including the amino end of the mature protein, was determined

C:Complex: monomer

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase; venom; zymogen

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-24/Domain: activation peptide #status predicted <ACP>

F:25-257/Product: venom B #status predicted <MAT>

F:25-243/Domain: trypsin homology <TRY>

F:31-162,49-65,97-255,141-209,173-188,199-224/Disulfide bonds: #status predicted

F:64,109,203/Active site: His, Asp, Ser #status predicted

Query Match 74.5%; Score 35; DB 1; Length 257;

Best Local Similarity 75.0%; Pred. No. 44;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I:|:|:|:|

Db 53 LLNEEWV 60

RESULT 26

I36945

haptoglobin Hp - chimpanzee (fragment)

C:Species: Pan troglodytes (chimpanzee)

C:Date: 16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change 22-Jun-1999

C:Accession: I36945

R:McEvoy, S.; Maeda, N.

J. Biol. Chem. 263, 15740-15747, 1988

A:Title: Complex events in the evolution of the haptoglobin gene cluster in primates.

A:Reference number: I36945; MUID:89008487

A:Accession: I36945

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 <RES>

A:Cross-references: GB:M20760; GB:J04045; NID:gl76788; PIDN:AAA35412.1; PID:gl76792

C:Genetics:

A:Gene: Hp

C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology

F:14-251/Domain: trypsin homology <TRY>

Query Match 74.5%; Score 35; DB 2; Length 258;

Best Local Similarity 62.5%; Pred. No. 44;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I:|:|:|:|

Db 43 LINEQWLL 50

RESULT 27

I36947

haptoglobin Hpp - chimpanzee (fragment)

C:Species: Pan troglodytes (chimpanzee)

C:Date: 16-Feb-1996 #sequence_revision 16-Feb-1996 #text_change 22-Jun-1999

C:Accession: I36947

R:McEvoy, S.; Maeda, N.

J. Biol. Chem. 263, 15740-15747, 1988

A:Title: Complex events in the evolution of the haptoglobin gene cluster in primates.

A:Reference number: I36945; MUID:89008487

A:Accession: I36947

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-258 <RES>

A:Cross-references: GB:M20762; GB:J04045; NID:gl76790; PIDN:AAA35414.1; PID:gl76794

C:Genetics:

A:Gene: Hpp

C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology

F:14-251/Domain: trypsin homology <TRY>

Query Match 74.5%; Score 35; DB 2; Length 258;

Best Local Similarity 62.5%; Pred. No. 44;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I:|:|:|:|

Db 43 LINEQWLL 50

RESULT 28

G02959

haptoglobin - rhesus macaque (fragment)

C:Species: Macaca mulatta (rhesus macaque)

C:Date: 21-Dec-1996 #sequence_revision 06-Jun-1997 #text_change 22-Jun-1999

C:Accession: G02959

R:Erickson, L.M.

submitted to the EMBL Data Library, September 1993

A:Reference number: G12739

A:Accession: G02959

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-258 <ERI>

A:Cross-references: EMBL:U02000; NID:g406874; PIDN:AAA17367.1; PID:g466455

C:Genetics:

A:Gene: HP
C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology
F:14-251/Domain: trypsin homology <TRY>

Query Match 74.5%; Score 35; DB 2; Length 258;
Best Local Similarity 62.5%; Pred. No. 44;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I::I::I::I

Db 43 LINEQWLL 50

RESULT 29

I47078

coagulation factor IXa (EC 3.4.21.22) - sheep (fragment)

C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)

C:Date: 15-Oct-1996 #sequence_revision 15-Oct-1996 #text_change 22-Jun-1999

C:Accession: I47078

R:Sarkar, G.; Koerberl, D.D.; Sommer, S.S.

Genomics 6, 133-143, 1990

A:Title: Direct sequencing of the activation peptide and the catalytic domain of the fac

A:Reference number: I46580; MUID:90152675

A:Accession: I47078

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-274 <SAR>

A:Cross-references: GB:M26233; NID:q165878; PIDN:AAA31520.1; PID:g552419

C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homology

C:Keywords: hydrolase; serine proteinase

F:49-274/Domain: trypsin homology (fragment) <TRY>

Query Match 74.5%; Score 35; DB 2; Length 274;

Best Local Similarity 62.5%; Pred. No. 47;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I::I::I::I

Db 78 IVNEKVV 85

RESULT 30

E70472

ribosomal protein S02 - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 13-Aug-1999

C:Accession: E70472

R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

V. Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: E70472

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-274 <AQ>

A:Cross-references: GB:AE000767; NID:g2984235; PIDN:AA007767.1; PID:g2984239; GB:AE00065

A:Experimental source: strain VF5

C:Genetics:

C:Gene: rpsB

C:Superfamily: Escherichia coli ribosomal protein S2

Query Match 74.5%; Score 35; DB 2; Length 274;

Best Local Similarity 100.0%; Pred. No. 47;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 VNERWV 8

I::I::I::I

Db 92 VNERWV 97

RESULT 31

A83757

hypothetical protein BH0857 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: A83757

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: A83757

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-282 <STO>

A:Cross-references: GB:AP001510; GB:BA000004; NID:gi0173440; PIDN:BA004576.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH0857

Query Match 74.5%; Score 35; DB 2; Length 282;

Best Local Similarity 62.5%; Pred. No. 48;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWV 8

I::I::I::I

Db 115 LVNERWI 122

RESULT 32

HPDG

haptoglobin precursor - dog (fragments)

C:Species: Canis lupus familiaris (dog)

C:Date: 09-Oct-1992 #sequence_revision 31-May-1996 #text_change 11-May-2000

C:Accession: JX0223; JX0224; B26503

R:Kunazaki, T.; Urushibara, N.; Ishii, S.

J. Biochem. 112, 11-19, 1992

A:Title: Amino acid sequence and disulfide-bridge location of canine haptoglobin.

A:Reference number: JX0223; MUID:93054403

A:Accession: JX0223

A:Molecule type: protein

A:Residues: 1-83 <KU2>

A:Experimental source: plasma

A>Note: Arg-84 was found at the end of some alpha chains; we have inserted it by homo

A:Accession: JX0224

A:Molecule type: protein

A:Residues: 85-329 <KU3>

A:Experimental source: plasma

R:Kurosky, A.; Kim, H.H.; Touchstone, B.

Comp. Biochem. Physiol. B 55, 453-459, 1976

A:Title: Comparative sequence analysis of the N-terminal region of rat, rabbit, and d

A:Reference number: A90931; MUID:77025019

A:Accession: B26503

A:Molecule type: protein

A:Residues: 85-124 <KUR>

C:Comment: The mature haptoglobin molecule is a dimer of heterodimers, the two chains

C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology

C:Keywords: acute phase; glycoprotein; hemoglobin binding; heterotetramer; iron trans

F:1-83/Product: haptoglobin alpha chain #status experimental <ALP>

F:15-68/Domain: complement factor H repeat homology <FH2>

F:85-329/Product: haptoglobin beta chain #status experimental <BET>

F:85-322/Domain: trypsin homology <TRY>

F:9,107,214/Binding site: carbohydrate

F:34-68,72-189,232-263,274-304/Disulfide bonds: #status experimental

Query Match 74.5%; Score 35; DB 1; Length 329;

Best Local Similarity 62.5%; Pred. No. 57;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I::I::I::I

Db 114 LINEQWLL 121

RESULT 33

I36941

haptoglobin - chimpanzee (fragment)
C:Species: Pan troglodytes (chimpanzee)
C:Date: 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change 22-Jun-1999
C:Accession: I36941
R:Brickson, L.M.; Kim, H.S.; Maeda, N.
Genomics 14, 948-958, 1992
A:Title: Junctions between genes in the haptoglobin gene cluster of primates.
A:Reference number: I36941; MUID:93122805

A:Accession: I36941

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-345 <RES>

A:Cross-references: GB:M84462; NID:gl76781; PIDN:AAA70196.1; PID:gl76784

C:Genetics:

A:Gene: Hp

A:Introns: 28/1; 62/1; 87/1

C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology

F:31-84/Domain: complement factor H repeat homology <FH2>

F:101-338/Domain: trypsin homology <TRY>

Query Match 74.5%; Score 35; DB 2; Length 345;

Best Local Similarity 62.5%; Pred. No. 59;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I::I::I::I

Db 130 LINEQWLL 137

RESULT 34

I36942

haptoglobin - chimpanzee (fragment)

C:Species: Pan troglodytes (chimpanzee)

C:Date: 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change 22-Jun-1999

C:Accession: I36942

R:Brickson, L.M.; Kim, H.S.; Maeda, N.

Genomics 14, 948-958, 1992

A:Title: Junctions between genes in the haptoglobin gene cluster of primates.

A:Reference number: I36941; MUID:93122805

A:Accession: I36942

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-346 <RES>

A:Cross-references: GB:M84463; NID:g903708; PIDN:AAA70197.1; PID:g903710

C:Genetics:

A:Gene: HPR

A:Introns: 29/1; 63/1; 88/1

C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology

F:32-85/Domain: complement factor H repeat homology <FH2>

F:102-339/Domain: trypsin homology <TRY>

Query Match

Best Local Similarity 74.5%; Score 35; DB 2; Length 346;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I::I::I::I

Db 131 LINEQWLL 138

RESULT 35

HPH01

haptoglobin precursor, allele 1 [validated] - human

N:Contains: haptoglobin alpha chain; haptoglobin beta chain

C:Species: Homo sapiens (man)

C:Date: 19-Feb-1984 #sequence_revision 30-Sep-1987 #text_change 08-Dec-2000

C:Accession: A93521; A91321; A94703; A91004; A00917; A22704; A22864; A25739

R:Brune, J.L.; Yang, F.; Barnett, D.R.; Bowman, B.H.

Nucleic Acids Res. 12, 4531-4538, 1984

A:Title: Evolution of haptoglobin: comparison of complementary DNA encoding Hpa1pha(1S)

A:Reference number: A93521; MUID:84247319

A:Accession: A93521

A:Molecule type: mRNA

A:Residues: 1-347 <BRU>

A:Cross-references: EMBL:X60037

R:van der Straten, A.; Herzog, A.; Cabazon, T.; Bollen, A.

FEBS Lett. 168, 103-107, 1984

A:Title: Characterization of human haptoglobin cDNAs coding for alpha-2FS-beta and alpha-2FS-beta

A:Reference number: A91321; MUID:84158941

A:Accession: A91321

A:Molecule type: mRNA

A:Residues: 1-347 <VAN>

A:Cross-references: EMBL:K01763; NID:g184316; PIDN:AAA52684.1; PID:g306880

R:Kurosky, A.; Barnett, D.R.; Lee, T.H.; Touchstone, B.; Hay, R.E.; Arnott, M.S.; Bo-

proc. Natl. Acad. Sci. U.S.A. 77, 3388-3392, 1980

A:Title: Covalent structure of human haptoglobin: a serine protease homolog.

A:Reference number: A94703; MUID:81013890

A:Contents: variant alpha-1S (slow); disulfide bonds

A:Accession: A94703

A:Molecule type: protein

A:Residues: 19-101;103-307,'N',309-320,'CS',323-337,'N',339-347 <KUR>

R:Bensi, G.; Raugel, G.; Klefenz, H.; Cortese, R.

EMBO J. 4, 119-126, 1985

A:Title: Structure and expression of the human haptoglobin locus.

A:Reference number: A91004; MUID:85257429

A:Contents: variant alpha-1F (fast)

A:Accession: A91004

A:Molecule type: DNA

A:Residues: 3-69,'DK',72-347 <BEN>

A:Cross-references: GB:X01793

R:Kurosky, A.; Kim, H.H.; Touchstone, B.

Comp. Biochem. Physiol. B 55, 453-459, 1976

A:Title: Comparative sequence analysis of the N-terminal region of rat, rabbit, and

A:Reference number: A90931; MUID:77025019

A:Contents: annotation; carbohydrate bound to beta chain

R:Malchty, B.; Dixon, G.H.

Can. J. Biochem. 51, 249-264, 1973

A:Title: Studies on the interchain disulfides of human haptoglobins.

A:Reference number: A90747; MUID:73166302

A:Contents: annotation; disulfide bonds in alpha chain

C:Comment: The sequence of the 1S variant is shown. There is a separate entry for h

C:Comment: The mature haptoglobin molecule is a dimer of heterodimers, the two chain

C:Comment: Each haptoglobin beta chain can bind an alpha-beta heterodimer of hemogl

C:Comment: The haptoglobin-hemoglobin complex exceeds the kidney threshold for excre

C:Comment: The haptoglobin-hemoglobin complex is degraded in the liver (where haptogl

C:Genetics:

A:Gene: GDB:HP

A:Cross-references: GDB:l19314; OMIM:l40100

A:Map position: 16q22.2-16q22.2

A:Introns: 2/2; 30/1; 64/1; 89/1

C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology

C:Keywords: acute phase; glycoprotein; hemoglobin binding; heterotetramer; iron tran-

F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-101/Product: haptoglobin alpha chain #status experimental <ACH>

F:33-86/Domain: complement factor H repeat homology <PH2>

F:103-347/Product: haptoglobin beta chain #status experimental <BCH>

F:103-340/Domain: trypsin homology <TRY>

F:33/Disulfide bonds: interchain #status experimental

F:52-86,90-207,250-281,292-322/Disulfide bonds: #status experimental

F:125,148,152,182/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 74.5%; Score 35; DB 1; Length 347;

Best Local Similarity 62.5%; Pred. No. 60;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

I::I::I::I

Db 132 LINEQWLL 139

RESULT 36

G00006

haptoglobin - black-handed spider monkey

C:Species: Ateles geoffroyi (black-handed spider monkey)

C:Date: 13-Mar-1997 #sequence_revision 13-Mar-1997 #text_change 22-Jun-1999
C:Accession: G00006
R:Erickson, L.M.
submitted to the EMBL Data Library, January 1994
A:Reference number: G00006
A:Accession: G00006
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-347 <ERI>
A:Cross-references: EMBL:U04852; NID:g440241; PIDN:AAA03727.1; PID:g440242
C:Genetics:
A:Introns: 2/2; 30/1; 64/1; 89/1
C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology
F:33-86/Domain: complement factor H repeat homology <FH2>
F:103-340/Domain: trypsin homology <TRY>

Query Match 74.5%; Score 35; DB 2; Length 347;
Best Local Similarity 62.5%; Pred. No. 60;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I:|:|:|:|
Db 132 LINEQWLL 139

RESULT 37

HPHUR

haptoglobin-related protein precursor - human

C:Species: Homo sapiens (man)

C:Date: 27-Nov-1985 #sequence_revision 27-Nov-1985 #text_change 18-Jun-1999

C:Accession: A00919; A30360; I61855; I60126

R:Maeda, N.

J. Biol. Chem. 260, 6698-6709, 1985

A:Title: Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair. The haptoglobin-related gene pair.

A:Reference number: A92532; MUID:85207676

A:Accession: A00919

A:Molecule type: DNA

A:Residues: 1-348 <MAE>

A:Cross-references: GB:M10935

R:Kuhajda, F.P.; Katumuluwa, A.I.; Pasternack, G.R.

Proc. Natl. Acad. Sci. U.S.A. 86, 1188-1192, 1989

A:Title: Expression of haptoglobin-related protein and its potential role as a tumor and

A:Reference number: A30360; MUID:89145208

A:Accession: A30360

A:Molecule type: protein

A:Residues: 68-73,'E',75:92-99,'Q',101-103 <KUH>

A:Note: the amino-terminal end of the mature alpha chain was found to be blocked

R:Erickson, L.M.; Kim, H.S.; Maeda, N.

Genomics 14, 948-958, 1992

A:Title: Junctions between genes in the haptoglobin gene cluster of primates.

A:Reference number: I36941; MUID:93122805

A:Accession: I61855

A:Status: translation not shown; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-348 <RES>

A:Cross-references: GB:M69197; NID:g292156; PIDN:AAA88079.1; PID:g292158

R:Maeda, N.; McEvoy, S.M.; Harris, H.F.; Huismann, T.H.; Smithies, O.

Proc. Natl. Acad. Sci. U.S.A. 83, 7395-7399, 1986

A:Title: Polymorphisms in the human haptoglobin gene cluster: chromosomes with multiple

A:Reference number: I60126; MUID:87016954

A:Accession: I60126

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 258-338,'H',340-348 <RES>

A:Cross-references: GB:M13908; NID:g184320; PIDN:AAA52686.1; PID:g184321

A:Experimental source: an individual (C.G.) with multiple haptoglobin-related genes

C:Comment: This protein is expressed during pregnancy and in some breast carcinomas. It

C:Genetics:

A:Gene: GDB:HPHUR

A:Cross-references: GDB:I19316; OMIM:140210

A:Map position: 16q22.1-16q22.1

A:Introns: 2/3; 31/1; 65/1; 90/1

A:Note: humans are polymorphic in the number of Hpr genes

C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology
C:Keywords: acute phase; blocked amino end; glycoprotein; hemoglobin binding; heterot
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-102/Product: haptoglobin-related protein alpha chain #status predicted <ALP>
F:34-87/Domain: complement factor H repeat homology <FH2>
F:104-348/Product: haptoglobin-related protein beta chain #status predicted <BET>
F:104-341/Domain: trypsin homology <TRY>
F:20/Modified site: blocked amino end (Leu) (in mature form) #status experimental
F:53-87,91-208,251-289,293-323/Disulfide bonds: #status predicted
F:126,149,153/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 74.5%; Score 35; DB 1; Length 348;
Best Local Similarity 62.5%; Pred. No. 60;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I:|:|:|:|
Db 133 LINEQWLL 140

RESULT 38

I36944

haptoglobin - chimpanzee (fragment)

C:Species: Pan troglodytes (chimpanzee)

C:Date: 04-Oct-1996 #sequence_revision 04-Oct-1996 #text_change 22-Jun-1999

C:Accession: I36944

R:Erickson, L.M.; Kim, H.S.; Maeda, N.

Genomics 14, 948-958, 1992

A:Title: Junctions between genes in the haptoglobin gene cluster of primates.

A:Reference number: I36941; MUID:93122805

A:Accession: I36944

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-349 <RES>

A:Cross-references: GB:M84463; NID:g903708; PIDN:AAA70198.1; PID:g304430

C:Genetics:

A:Gene: HPP

A:Introns: 3/1; 32/1; 66/1; 91/1

C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology

F:35-88/Domain: complement factor H repeat homology <FH2>

F:105-342/Domain: trypsin homology <TRY>

Query Match 74.5%; Score 35; DB 2; Length 349;
Best Local Similarity 62.5%; Pred. No. 60;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I:|:|:|:|
Db 134 LINEQWLL 141

RESULT 39

HPHUR2

haptoglobin precursor, allele 2 [validated] - human

N:Alternate names: haptoglobin-2

C:Species: Homo sapiens (man)

C:Date: 19-Feb-1984 #sequence_revision 27-Nov-1985 #text_change 08-Dec-2000

C:Accession: A92532; B25739; A93973; A93485; I61854; A00918

R:Maeda, N.

J. Biol. Chem. 260, 6698-6709, 1985

A:Title: Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair. The haptoglobin-related gene pair.

A:Reference number: A92532; MUID:85207676

A:Accession: A92532

A:Molecule type: DNA

A:Residues: 1-406 <MAE>

A:Cross-references: GB:M10935; NID:g184327; PIDN:AAA88080.1; PID:g386783

R:van der Straten, A.; Herzog, T.; Bollen, A.

FEBS Lett. 168, 103-107, 1984

A:Title: Characterization of human haptoglobin cDNAs coding for alpha-2FS-beta and a

A:Reference number: A91321; MUID:84158941

A:Accession: B25739

A:Molecule type: mRNA

A:Residues: 1-406 <VAN>

A:Cross-references: EMBL:K01763
R:Yang, F.; Brune, J.L.; Baldwin, W.D.; Barnett, D.R.; Bowman, B.H.
Proc. Natl. Acad. Sci. U.S.A. 80, 5875-5879, 1983
A:Title: Identification and characterization of human haptoglobin cDNA.
A:Reference number: A93973; MUID:83300066
A:Accession: A93973
A:Molecule type: mRNA
A:Residues: 1-69, 'N', 71-406 <YAN>
A:Cross-references: EMBL:K00422; NID:g184322; PIDN:AAA52687.1; PID:g306882
R:Rauegi, G.; Bensi, G.; Colantuoni, V.; Romano, V.; Santoro, C.; Costanzo, F.; Cortese, Nucleic Acids Res. 11, 5811-5819, 1983
A:Title: Sequence of human haptoglobin cDNA: evidence that the alpha and beta subunits are
A:Reference number: A93485; MUID:83399252
A:Accession: A93485
A:Molecule type: mRNA
A:Residues: 3-366, 'N', 368-406 <RAU>
A:Cross-references: GB:M12387
R:Kurosky, A.; Barnett, D.R.; Lee, T.H.; Touchstone, B.; Hay, R.E.; Arnott, M.S.; Bowman Proc. Natl. Acad. Sci. U.S.A. 77, 3388-3392, 1980
A:Title: Covalent structure of human haptoglobin: a serine protease homolog.
A:Reference number: A94703; MUID:81013890
A:Contents: annotation; revisions to the alpha-2 chain deduced from comparisons with the
R:Malchy, B.; Dixon, G.H.
Can. J. Biochem. 51, 249-264, 1973
A:Title: Studies on the interchain disulfides of human haptoglobins.
A:Reference number: A90747; MUID:73166302
A:Contents: annotation; disulfide bonds
R:Erickson, L.M.; Kim, H.S.; Maeda, N. Genomics 14, 948-958, 1992
A:Title: Junctions between genes in the haptoglobin gene cluster of primates.
A:Reference number: I36941; MUID:9312805
A:Accession: I61854
A>Status: translation not shown; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-406 <RES>
A:Cross-references: GB:M69197; NID:g292156; PIDN:AAA88078.1; PID:g292157
C:Comment: This haptoglobin with a partial duplication in the alpha chain region is a co
C:Genetics:
A:Gene: GDB:HP
A:Cross-references: GDB:119314; OMIM:140100
A:Map position: 16q22.2-16q22.2
A:Introns: 2/2; 30/1; 64/1; 89/1; 123/1; 148/1
C:Superfamily: haptoglobin; complement factor H repeat homology; trypsin homology
C:Keywords: acute phase; duplication; glycoprotein; hemoglobin binding; heterotetramer;
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-160/Product: haptoglobin allele 2 alpha chain #status experimental <ACH>
F:33-86/Domain: complement factor H repeat homology #status atypical <FHL>
F:92-145/Domain: complement factor H repeat homology #status predicted <BCH>
F:162-406/Product: haptoglobin beta chain #status predicted <TRY>
F:162-398/Domain: trypsin homology
F:33-92/Disulfide bonds: interchain #status experimental
F:52-86,111-145/Disulfide bonds: #status experimental
F:184, 207, 211, 241/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:309-340,351-381/Disulfide bonds: #status predicted

Query Match 74.5%; Score 35; DB 1; Length 406;
Best Local Similarity 62.5%; Pred. No. 71;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERVL 9
I:|:|:|
DB 191 LINEQWL 198

RESULT 40
T46034
branched-chain-amino-acid transaminase-like protein - Arabidopsis thaliana
N:Alternate names: protein T16K5.30
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 02-Sep-2000
C:Accession: T46034
R:Rieger, M.; Gabel, C.; Mueller-Auer, S.; Schaefer, M.; Zipp, M.; Mewes, H.W.; Lemcke, submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23015
A:Accession: T46034
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-410 <RIE>
A:Cross-references: EMBL:AL132965
A:Experimental source: cultivar Columbia; BAC clone T16K5
C:Genetics:
A:Map position: 3
A:Introns: 42/3; 70/2; 129/3; 189/3; 233/3; 266/3; 302/3; 334/3; 372/2
A:Note: T16K5.30
C:Superfamily: branched-chain-amino-acid transaminase BAT1

Query Match 74.5%; Score 35; DB 2; Length 410;
Best Local Similarity 75.0%; Pred. No. 71;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWV 8
I:|:|:|
DB 183 VLANKRWV 190

RESULT 41
KFBO
coagulation factor IXa (EC 3.4.21.22) precursor - bovine
N:Alternate names: Christmas factor
C:Species: Bos primigenius taurus (cattle)
C:Date: 30-Nov-1980 #sequence_revision 03-Aug-1984 #text_change 16-Jul-1999
R:Katsuyama, K.; Ericsson, L.H.; Enfield, D.L.; Walsh, K.A.; Neurath, H.; Davie, E.W., Proc. Natl. Acad. Sci. U.S.A. 76, 4990-4994, 1979
A:Title: Comparison of amino acid sequence of bovine coagulation factor IX (Christmas
A:Reference number: A14757; MUID:80056619
A:Accession: A14757
A:Molecule type: protein
A:Residues: 1-63, 'T', 65-416 <KAT>
R:McMullen, B.A.; Fujikawa, K.; Kisiel, W. Biochem. Biophys. Res. Commun. 115, 8-14, 1983
A:Title: The occurrence of beta-hydroxyaspartic acid in the vitamin K-dependent blood
A:Reference number: A20274; MUID:83308813
A:Accession: B20274
A:Molecule type: protein
A:Residues: 59-63, 'X', 65-69 <MCM>
R:Choo, K.H.; Gould, K.G.; Rees, D.J.G.; Brownlee, G.G. Nature 299, 178-180, 1982
A:Title: Molecular cloning of the gene for human anti-haemophilic factor IX.
A:Reference number: I45891; MUID:82272386
A:Accession: I45891
A>Status: translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 52-139 <CHO>
A:Cross-references: GB:J00007; NID:g163053; PIDN:AAA30520.1; PID:g163054
R:Hase, S.; Kawabata, S.; Nishimura, H.; Takeya, H.; Sueyoshi, T.; Miyata, T.; Iwanay J. Biochem. 104, 867-868, 1988
A:Title: A new trisaccharide sugar chain linked to a serine residue in bovine blood
A:Reference number: A44556; MUID:89213999
A:Accession: A44556
A:Contents: annotation
C:Note: structure and location of a carbohydrate covalently bound to Ser
C:Comment: Factor IX is activated by factor Xla, which excises the activation peptide
C:Comment: The gamma-carboxyglutamic acid residues arise by posttranslational, vitam
C:Comment: Calcium binds to the gamma-carboxyglutamic acid (Gla) residues and, with s
C:Function:
A:Description: catalyzes the proteolytic activation of coagulation factor X in the pi
A:Pathway: blood coagulation intrinsic pathway
C:Superfamily: coagulation factor X; EGF homology; Gla domain homology; trypsin homol
C:Keywords: beta-hydroxyaspartic acid; blood coagulation; calcium binding; carboxyglu
F:1-146/Product: coagulation factor IXa light chain #status experimental <ALC>
F:1-45/Domain: Gla domain homology (fragment) <GLA>
F:51-82/Domain: EGF homology <EG1>
F:88-124/Domain: EGF homology <EG2>
F:147-181/Domain: activation peptide
F:182-416/Product: coagulation factor IXa heavy chain #status experimental <AHC>
F:182-409/Domain: trypsin homology <TRY>

F:7,8,15,17,20,21,26,27,30,33,36,40/Modified site: gamma-carboxyglutamic acid (Glu) #sta
 F:18-23,51-62,56-71,73-82,88-99,95-109,111-124,132-290,207-223,337-351,362-390/Disulfide
 F:53/Binding site: carboxylate (Ser) (covalent) #status experimental
 F:64/Modified site: erythro-beta-hydroxyaspartic acid (Asp) #status experimental
 F:158,168,173,261/Binding site: carboxylate (Asn) (covalent) #status experimental
 F:222,270,366/Active site: His, Asp, Ser #status predicted

Query Match 74.5%; Score 35; DB 1; Length 416;
 Best Local Similarity 62.5%; Pred. No. 73;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
 :|||:|:
 Db 211 IVNEKRVV 218

RESULT 42
 S11674
 acrosin (BC 3.4.21.10) precursor - human
 C:Species: Homo sapiens (man)
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000
 C:Accession: S11674; S23499; S12063; A61022; S03330
 R:Keime, S.; Adham, I.M.; Engel, W.
 Eur. J. Biochem. 190, 195-200, 1990

A:Title: Nucleotide sequence and exon-intron organization of the human proacrosin gene.
 A:Reference number: S11674; MUID:90306003
 A:Accession: S11674
 A:Molecule type: DNA
 A:Residues: 1-421 <KEI>
 A:Cross-references: EMBL:X54017; NID:g35582; PIDN:CAA37964.1; PID:g1216165
 A:Note: the authors translated the codon AGG for residue 64 as Thr and CTG for residue 2
 R:Vazquez-Levin, M.H.; Reventosa, J.; Gordon, J.W.
 Eur. J. Biochem. 207, 23-26, 1992

A:Title: Molecular cloning, sequencing and restriction mapping of the genomic sequence
 A:Reference number: S23499; MUID:92331659
 A:Accession: S23499
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-421 <VAZ>
 A:Cross-references: EMBL:M77378
 A:Note: the nucleotide sequence was submitted to the EMBL Data Library, October 1992
 R:Keime, S.
 submitted to the EMBL Data Library, December 1989
 A:Reference number: S12063
 A:Accession: S12063
 A:Molecule type: DNA
 A:Residues: 1-225, 'R', 227-421 <KEI2>
 A:Cross-references: EMBL:X54017
 R:Adham, I.M.; Klemm, U.; Maier, W.M.; Engel, W.
 Hum. Genet. 84, 125-128, 1990

A:Title: Molecular cloning of human proacrosin cDNA.
 A:Reference number: A61022; MUID:90128988
 A:Accession: A61022
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-225, 'V', 227-267, 'R', 269-421 <ADR>
 R:Baba, T.; Watanabe, K.; Kashiwabara, S.I.; Arai, Y.
 FEBS Lett. 244, 296-300, 1989

A:Title: Primary structure of human proacrosin deduced from its cDNA sequence.
 A:Reference number: S03330; MUID:89153568
 A:Accession: S03330
 A:Molecule type: mRNA
 A:Residues: 1-63, 'T', 65-119, 'V', 121-165, 'L', 167-267, 'R', 269-344, 'R', 346-421 <BAB>
 A:Cross-references: EMBL:Y00970; NID:g28325; PIDN:CAA68784.1; PID:g28326
 C:Genetics:
 A:Gene: GDB:ACR
 A:Cross-references: GDB:119645; OMIM:102480
 A:Map position: 22q13-22qter
 A:Introns: 26/2; 94/2; 185/2; 237/3
 C:Superfamily: acrosin; trypsin homology
 C:Keywords: glycoprotein; hydrolase; serine proteinase; sperm
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-42/Product: acrosin #status predicted <NAT>

F:20-42/Product: acrosin light chain #status predicted <LCH>
 F:43-421/Product: acrosin heavy chain #status predicted <HCH>
 F:43-285/Domain: trypsin homology <TRY>
 F:302-379/Region: proline-rich
 F:222/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:25-154/Disulfide bonds: #status predicted
 F:29-162/Disulfide bonds: #status predicted
 F:73-89/Disulfide bonds: #status predicted
 F:88,142,240/Active site: His, Asp, Ser #status predicted
 F:177-246/Disulfide bonds: #status predicted
 F:209-225/Disulfide bonds: #status predicted
 F:236-266/Disulfide bonds: #status predicted

Query Match 74.5%; Score 35; DB 1; Length 421;
 Best Local Similarity 75.0%; Pred. No. 73;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
 :|||:|:
 Db 77 LLNSRWVL 84

RESULT 43
 G96695
 hypothetical protein F5A8.9 [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: G96695
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: G96695
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-453 <STO>
 A:Cross-references: GB:AE005173; NID:g4204282; PIDN:AAD10663.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: F5A8.9
 A:Map position: 1

Query Match 74.5%; Score 35; DB 2; Length 453;
 Best Local Similarity 75.0%; Pred. No. 79;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWV 8
 :|||:|:
 Db 60 VVYNHRWV 67

RESULT 44
 C86443
 unknown protein [imported] - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
 C:Accession: C86443
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 ansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

Search completed: November 6, 2002, 12:08:14
Job time : 13 secs

A:Reference number: A86141; MUID:21016719
A:Accession: C86443
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-505 <STO>
A:Cross-references: GB:AE005172; NID:g11136726; PIDN:AG31307.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 74.5%; Score 35; DB 2; Length 505;
Best Local Similarity 66.7%; Pred. No. 89;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||:|:|:|
DB 371 VLTSEKWL 379

RESULT 45
S54045
probable membrane protein YDR061w - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein D4241; hypothetical protein YBR539; hypothetical
C:Species: Saccharomyces cerevisiae
C>Date: 08-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 02-Feb-2001
C:Accession: S54045; S58841; S61750; S67877
R:Hunt, S.; Bowman, S.; Harris, D.
submitted to the EMBL Data Library, May 1995
A:Reference number: S54031
A:Accession: S54045
A:Molecule type: DNA
A:Residues: 1-539 <HUN>
A:Cross-references: EMBL:249209; NID:g798897; PIDN:CAA89090.1; PID:g798912
R:Brandt, P.; Otto, B.; Ramlow, S.; Blocker, H.
submitted to the EMBL Data Library, January 1995
A:Reference number: S58832
A:Accession: S58841
A:Molecule type: DNA
A:Residues: 1-539 <BRA>
A:Cross-references: EMBL:X84162; NID:g706817; PIDN:CAA58977.1; PID:g706827
R:Brandt, P.; Ramlow, S.; Otto, B.; Blocker, H.
Yeast 12, 85-90, 1996
A:Title: Nucleotide sequence analysis of a 32,500 bp region of the right arm of Sacchar
A:Reference number: S61741; MUID:96381250
A:Accession: S61750
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-539 <BRW>
A:Cross-references: EMBL:X84162; NID:g706817; PIDN:CAA58977.1; PID:g706827
R:Blocker, H.; Brandt, P.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S67587
A:Accession: S67877
A:Molecule type: DNA
A:Residues: 1-539 <BLO>
A:Cross-references: EMBL:274357; NID:g1431510; PIDN:CAA98879.1; PID:g1431511; MIPS:YDR06
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 4R
C:Superfamily: yeast probable membrane protein YDR061w; ATP-binding cassette homology
C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
F:322-530/Domain: ATP-binding cassette homology <ABC>
F:339-346/Region: nucleotide-binding motif A (P-loop)
F:463-481/Domain: transmembrane #status predicted <TMM>

Query Match 74.5%; Score 35; DB 2; Length 539;
Best Local Similarity 55.6%; Pred. No. 96;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
:|:|:|:
DB 36 ILPNEKWI 44

THIS PAGE BLANK (USPIC)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run On: November 6, 2002, 12:01:15 ; Search time 23.7778 Seconds
(without alignments)
42.042 Million cell updates/sec

Title: US-09-905-083-34
Perfect score: 47
Sequence: 1 VLVNERWVL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	47	100.0	9	AAE08239	Human stratum corn
2	47	100.0	136	ABG23378	Novel human diagno
3	47	100.0	225	AAB98502	Human Stratum Corn
4	47	100.0	253	AAR67888	Human Stratum corn
5	47	100.0	253	AAW05383	Human amyloid prec
6	47	100.0	257	AAAB21326	Human HSCSE. Homo
7	43	91.5	9	AAE08245	Human stratum corn
8	43	91.5	9	AAE08255	Human stratum corn
9	39	83.0	212	AAW28590	Human Factor D. H
10	39	83.0	250	AAW05421	Human adipsin/D en
11	38	80.9	64	AAW01196	Serine protease Pf

12	38	80.9	64	22	AAE08239	Flea serine protea
13	38	80.9	223	17	AAW01165	Serine protease Pf
14	38	80.9	224	22	AAE08239	Flea serine protea
15	38	80.9	232	19	AAW64261	Kallikrein substra
16	38	80.9	260	17	AAW10694	Human recombinant
17	38	80.9	260	18	AAW12393	Mouse neuropsin pr
18	37	78.7	233	19	AAW46773	Amino acid sequenc
19	37	78.7	233	22	AAW79000	Mamushi fibrinolyt
20	37	78.7	237	19	AAW41955	Flea serine protea
21	37	78.7	243	22	AAE08239	Drosophila melanog
22	37	78.7	258	19	AAW41954	Flea serine protea
23	37	78.7	258	19	AAW41954	Flea serine protea
24	37	78.7	288	22	AAU03901	Human protease-lik
25	37	78.7	348	22	AAU03900	Human protease-lik
26	37	78.7	1342	21	AAAG31251	Arabidopsis thalia
27	37	78.7	1382	21	AAAG31250	Arabidopsis thalia
28	37	78.7	1421	21	AAAG31249	Arabidopsis thalia
29	36	76.6	111	21	AAAB21302	Human KLK-L5 prote
30	36	76.6	162	20	AAAY28642	Human secreted pro
31	36	76.6	184	21	AAAB21301	Human KLK-L5 prote
32	36	76.6	238	22	ABG09697	Novel human diagno
33	36	76.6	248	21	AAAB21304	Human KLK-L5 prote
34	36	76.6	248	21	AAAB24032	Human PRO1303 prot
35	36	76.6	248	21	AAAY99393	Human PRO1303 prot
36	36	76.6	248	21	AAAY99393	Human PRO1303 (UNQ
37	36	76.6	248	22	AAAM23994	Human EST encoded
38	36	76.6	248	22	AAE66142	Protein of the inv
39	36	76.6	254	21	AAAB21303	Human KLK-L5 prote
40	36	76.6	258	15	AAAR55757	Serine protease RN
41	36	76.6	547	22	ABG63123	Drosophila melanog
42	36	76.6	580	22	ABG60547	Drosophila melanog
43	35	74.5	9	22	AAE08232	Human stratum corn
44	35	74.5	60	21	AAAY51621	Rabbit Factor IX p
45	35	74.5	60	21	AAAY51623	Bovine Factor IX p

ALIGNMENTS

RESULT 1

AAE08239
ID AAE08239 standard; peptide: 9 AA.

AC AAE08239;

XX 01-NOV-2001 (first entry)

DF Human stratum corneum chymotrypsin enzyme peptide #3 (residues 58-66).

DE Stratum corneum chymotrypsin enzyme; SCCE: cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antitense therapy; malignant hyperplasia.

XX Homo sapiens.

XX WO200159158-A1.

XX 16-AUG-2001.

PD 07-FEB-2001; 2001WO-US039977.

XX 11-FEB-2000; 2000US-0502600.

XX (UYAR-) UNIV ARKANSAS.

XX O'brien TJ;

XX WPI; 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -

PS Claim 25; Page 103; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.

XX Sequence 9 AA;
SQ Query Match 100.0%; Score 47; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
DB 1 VLVNERWVL 9

RESULT 2
ABG23378
ID ABG23378 standard; Protein: 136 AA.

XX
AC ABG23378;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #23369.

DE Human: chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS87565.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID No 53737; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 136 AA;

XX Query Match 100.0%; Score 47; DB 22; Length 136;
Best Local Similarity 100.0%; Pred. No. 0.61;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
DB 64 VLVNERWVL 72

RESULT 3
AAB98502
ID AAB98502 standard; Protein: 225 AA.

XX AAB98502;

XX 03-AUG-2001 (first entry)

XX Human Stratum Corneum Chymotryptic Enzyme, SCCE, catalytic domain.

XX Human: TADG-15; cytostatic; vaccine; ovarian tumour; cancer;

XX tumour antigen-derived gene 15; serine protease;

XX Stratum Corneum Chymotryptic Enzyme; SCCE.

XX Homo sapiens.

XX WO200129056-A1.

XX 26-APR-2001.

XX 20-OCT-2000; 2000WO-US29095.

XX 20-OCT-1999; 99US-0421213.

XX (UYAR-) UNIV ARKANSAS.

XX O'Brien TJ, Tanimoto H;

XX WPI; 2001-381031/40.

XX Novel extracellular serine protease, termed tumor antigen-derived gene
PT 15 protein overexpressed in carcinomas and DNA encoding it, for
PT diagnosis, treatment, prevention of cancer, particularly breast,
PT ovarian cancer.

XX Example 10; Fig 1; 130pp; English.

XX The present invention relates to human tumour antigen-derived gene 15
CC (TADG-15) protein and coding sequence (see AAH23601 and AAB98500).
CC TADG-15 is an extracellular serine protease. It was found that TADG-15 is
CC over-expressed in ovarian tumours. TADG-15 protein or its fragments of
CC 9-20 residues that lack TADG-15 protease activity are useful for
CC vaccinating an individual against TADG-15, having, suspected of having or
CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a
CC diagnostic or therapeutic target in cancer. The present sequence was used
CC in a sequence homology alignment with the catalytic domain of TADG-15.

XX Sequence 225 AA;

XX Query Match 100.0%; Score 47; DB 22; Length 225;
Best Local Similarity 100.0%; Pred. No. 1.1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9
 |||||
 Db 30 VLVNERWVL 38

RESULT 4

AAR67888
 ID AAR67888 standard; Protein; 253 AA.

AC AAR67888;
 XX

DT 09-AUG-1995 (first entry)
 XX

DE Human stratum corneum chymotryptic recombinant enzyme (SCCE).
 XX

KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
 KW callosities; keratosis pilaris; ichthyoses; eczema.

OS Homo sapiens.
 XX

PN WO9500651-A.
 XX

PD 05-JAN-1995.
 XX

PF 20-JUN-1994; 94WO-IB00166.
 XX

PR 18-JUN-1993; 93DK-0000725.
 XX

PA (SYMB-) SYMBICOM AB.
 XX

PI Egelrud T, Hansson L;
 XX

DR WPI: 1995-052088/07.
 DR N-PSDB; AAQ81203.

XX
 Nucleotide sequences encoding stratum corneum chymotryptic enzyme

PT - and related vectors, transformed cells and polypeptides,
 PT useful for treating skin disorders, e.g. acne or psoriasis, and
 PT for identification of specific inhibitors.

XX Disclosure; Page 97; 137pp; English.

XX The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
 CC and skin care products, especially to treat and prevent acne,
 CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
 CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
 CC produced recombinantly following mammal, insect, plant, or
 CC microorganism transformation with plasmid p507.

XX Sequence 253 AA;
 SQ

Query Match 100.0%; Score 47; DB 16; Length 253;
 Best Local Similarity 100.0%; Pred. NO. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9
 |||||
 Db 58 VLVNERWVL 66

RESULT 5

AAW05383
 ID AAW05383 standard; Protein; 253 AA.

AC AAW05383;
 XX

DT 31-DEC-1996 (first entry)
 XX

DE Human amyloid precursor protein protease.
 XX

KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;
 KW therapy.

XX

OS Homo sapiens.
 XX

PN WO9631122-A1.
 XX

PD 10-OCT-1996.
 XX

PF 02-APR-1996; 96WO-US04294.
 XX

PR 04-APR-1995; 95US-0416257.
 XX

PA (ELIL) LILLY & CO ELI.
 XX

PI Dixon EP, Johnstone EM, Little SP;
 XX

DR WPI: 1996-464694/46.
 DR N-PSDB; AAT39783.

XX New isolated human amyloid precursor protein protease - used to
 PT develop prods. for the treatment or diagnosis of associated
 PT conditions, esp. Alzheimer's disease

XX Claim 1; Page 44-45; 55pp; English.
 XX

CC Human amyloid precursor protein protease (AAW05383) is involved in
 CC the processing or clearance of amyloid precursor protein to form
 CC beta-amyloid peptide. Its amino acid sequence was deduced from
 CC a cDNA clone (AAT39783) obt'd. from a human lung library. Recombinant
 CC protease can be produced in transformed or transfected prokaryotic
 CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
 CC used to develop products for the design and testing of cpds. useful
 CC for treating or preventing conditions associated with beta-amyloid
 CC peptide, esp. Alzheimer's disease.

XX Sequence 253 AA;
 SQ

Query Match 100.0%; Score 47; DB 17; Length 253;
 Best Local Similarity 100.0%; Pred. NO. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLVNERWVL 9
 |||||
 Db 58 VLVNERWVL 66

RESULT 6

AAW05386
 ID AAW05386 standard; Protein; 257 AA.

XX
 AC AAB21326;
 XX

DT 02-FEB-2001 (first entry)
 XX

DE Human HSCEE.
 XX

KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCEE;
 KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;
 KW serine protease; cytostatic; cancer; prostate cancer.

XX Homo sapiens.
 OS

XX WO200053776-A2.
 PN

XX 14-SEP-2000.
 PD

XX 09-MAR-2000; 2000WO-CA00258.
 PF

XX 11-MAR-1999; 99US-0124260.
 PR

XX 01-APR-1999; 99US-0127386.
 PR

XX 21-JUL-1999; 99US-0144919.
 PR

XX (MOUN) MOUNT SINAI HOSPITAL.
 PA

XX Yousef GM, Diamandis EP;
 PI

XX WPI; 2000-587440/55.
 XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
 PT protein mediated disorders, especially cancer. -
 XX Example 4; Fig 17; 184pp; English.
 PS The present sequence is human stratum corneum chymotryptic enzyme
 CC (HSCEE), a member of the kallikrein multi-gene family. Kallikreins and
 CC kallikrein-like proteins are a subgroup of the serine protease enzyme
 CC family. They catalyse the selective cleavage of specific polypeptide
 CC precursors to release peptides with potent biological activity. Nucleic
 CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
 CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
 CC treatment, monitoring and diagnosis of cancers, especially prostate
 CC cancer. They can also be used to identify a substance that can associate
 CC with or mediate the biological activity of the proteins. Antibodies can
 CC be used to treat conditions mediated by the kallikrein-like proteins.
 XX Sequence 257 AA;
 SQ Query Match 100.0%; Score 47; DB 21; Length 257;
 Best Local Similarity 100.0%; Pred. No. 1.2;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VLVNERWVL 9
 Db 63 VLVNERWVL 71
 RESULT 7
 AAE08245
 ID AAE08245 standard; peptide; 9 AA.
 AC AAE08245;
 XX 01-NOV-2001 (first entry)
 DT Human stratum corneum chymotrypsin enzyme peptide #10 (residues 57-65).
 DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
 XX cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
 KW antisense therapy; malignant hyperplasia.
 XX Homo sapiens.
 OS WO200159158-A1.
 XX 16-AUG-2001.
 PD 07-FEB-2001; 2001WO-US03977.
 XX 11-FEB-2000; 2000US-0502600.
 PR (UYAR-) UNIV ARKANSAS.
 PA O'brien TJ;
 XX WPI; 2001-514676/56.
 DR Diagnosing cancer comprises detecting stratum corneum chymotrypsin
 PT enzyme -
 XX Disclosure; Page 104; 127pp; English.
 PS The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.
 XX Sequence 9 AA;
 SQ Query Match 91.5%; Score 43; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LVNERWVL 9
 Db 1 LVNERWVL 8
 RESULT 9
 AAY28590
 ID AAY28590 standard; Protein; 212 AA.

CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.
 XX Sequence 9 AA;
 SQ Query Match 91.5%; Score 43; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 VLVNERWV 8
 Db 2 VLVNERWV 9
 RESULT 8
 AAE08255
 ID AAE08255 standard; peptide; 9 AA.
 XX AAE08255;
 AC AAE08255;
 XX 01-NOV-2001 (first entry)
 DT Human stratum corneum chymotrypsin enzyme peptide #20 (residues 59-67).
 DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
 XX cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
 KW antisense therapy; malignant hyperplasia.
 XX Homo sapiens.
 OS WO200159158-A1.
 XX 16-AUG-2001.
 PD 07-FEB-2001; 2001WO-US03977.
 XX 11-FEB-2000; 2000US-0502600.
 PR (UYAR-) UNIV ARKANSAS.
 PA O'brien TJ;
 XX WPI; 2001-514676/56.
 DR Diagnosing cancer comprises detecting stratum corneum chymotrypsin
 PT enzyme -
 XX Disclosure; Page 106; 127pp; English.
 PS The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.
 XX Sequence 9 AA;
 SQ Query Match 91.5%; Score 43; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LVNERWVL 9
 Db 1 LVNERWVL 8
 RESULT 9
 AAY28590
 ID AAY28590 standard; Protein; 212 AA.

XX AAY28590;
AC 26-NOV-1999 (first entry)
XX Human Factor D.
DE
XX
KW Factor D; alternative complement pathway; serine protease; asthma;
KW cardiopulmonary bypass; tissue damage; Alzheimer's disease;
KW multiple sclerosis; acute myocardial infarction; Crohn's disease;
KW transplant rejection; systemic lupus erythematosus; inflammation.
XX
OS Homo sapiens.
XX
XX WO9942133-A1.
XX
XX 26-AUG-1999.
XX
XX 19-FEB-1999; 99WO-US03566.
XX
XX 20-FEB-1998; 98US-0075328.
XX
XX (TANO-) TANOX INC.
XX
XX Fung MSC, Sun BNC, Sun CRY;
XX
XX WPI: 1999-527424/44.
XX N-PSDB; AA206673.
XX
PT Factor D inhibitors useful for treating pathological inflammation
PT and autoimmune diseases
XX
XX Example 1; Page 93-94; 99pp; English.
XX
CC This sequence is human factor D. Factor D is a highly specific serine
CC protease essential for activation of the alternative complement pathway.
CC The plasma concentration of factor D in humans is very low making it a
CC suitable target for inhibition. It is the limiting enzyme for the
CC activation of the alternative complement pathway. Factor D is used to
CC generate inhibitors of complement activation which bind factor D at a
CC molar ratio of at least 1:5:1 and less than 80:1 (inhibitor:factor D).
CC Factor D was used to design a monoclonal antibody MAB 166-32 which binds
CC to factor D and blocks its ability to activate complement. The
CC anti-factor D molecules can function to inhibit in vivo complement
CC activation and/or the alternative complement pathway and inflammatory
CC events which occur. These include the recruitment and activation of
CC macrophages, neutrophils, platelets, and mast cells, oedema, and tissue
CC damage. The inhibitors of complement activation may be used to treat
CC diseases or conditions that are mediated by excessive or uncontrolled
CC activation of the complement system. These inhibitors are also
CC administered for treating complement-mediated conditions associated with
CC cardiopulmonary bypass. Diseases or conditions that are mediated by
CC excessive or uncontrolled activation of the complement systems include
CC tissue damage due to ischaemia-reperfusion following acute myocardial
CC infarction, aneurysm, stroke, haemorrhagic shock, crush injury, multiple
CC organ failure; inflammatory disorders, e.g. burns, endotoxaemia and
CC septic shock, haemodialysis, severe asthma, Crohn's disease; transplant
CC rejection; and adverse drug reaction. Autoimmune disorders can also be
CC treated such as systemic lupus erythematosus, Alzheimer's disease and
CC multiple sclerosis. The anti-factor D molecules can also be used
CC diagnostically to determine the presence of or quantity of factor D in a
CC tissue specimen or a body fluid sample.
XX
XX Sequence 212 AA;
SQ

Query Match 83.0%; Score 39; DB 20; Length 212;
Best Local Similarity 88.9%; Pred. No. 26;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERRWL 9
||| |||||
DB 29 VLVNERRWL 37

RESULT 10
AAR05421
ID AAR05421 standard; protein; 250 AA.
XX
XX AAR05421;
XX
XX 30-JUL-1990 (first entry)
XX
XX Human adipsin/D encoded by a cDNA.
XX
XX CAT; hybrid protein; Human adipsin/D.
XX
XX Homo sapiens.
XX
XX WO9001540-A.
XX
XX 22-FEB-1990.
XX
XX 09-AUG-1989; 89WO-US03417.
XX
XX 11-AUG-1988; 88US-0231224.
XX
XX (CALB-) CALIF BIOTECHN INC.
XX
XX Hilliker S, White R;
XX
XX WPI: 1990-083499/11.
XX N-PSDB; AAQ03566.
XX
XX Heterologous protein expression on prokaryotic host -
XX using 3' truncated chloramphenicol acetyl transferase gene to
XX stably express hybrid protein.
XX
XX Example; Fig 10; 67pp; English.
XX
XX When inserted into an expression vector, pTrpCAT 72,
XX the construct gave 10-15% levels of fusion protein upon
XX induction in W3110 cells.
XX See also AAQ03557 to AAQ05366; and AAQ04767.
XX
XX Sequence 250 AA;
SQ

Query Match 83.0%; Score 39; DB 11; Length 250;
Best Local Similarity 88.9%; Pred. No. 31;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERRWL 9
||| |||||
DB 51 VLVNERRWL 59

RESULT 11
AAW01196
ID AAW01196 standard; Peptide; 64 AA.
XX
XX AAW01196;
XX
XX 15-MAR-1997 (first entry)
XX
XX Serine protease PfSP17 N-terminal fragment.
XX
XX Flea; midgut; serine protease; PfSP17; vaccine.
XX domestic animal; infestation; insecticide; protease-inhibitor;
XX controlled release formulation; synergist.
XX
XX Siphonaptera sp.
XX
XX WO9611706-A1.
XX
XX 25-APR-1996.
XX
XX 18-OCT-1995; 95WO-US14442.
XX
XX

```

XX PR 07-JUN-1995; 95US-0485455.
XX PR 18-OCT-1994; 94US-0326773.
XX PR 07-JUN-1995; 95US-0482130.
XX PR 07-JUN-1995; 95US-0484211.
XX PR 07-JUN-1995; 95US-0485443.
XX PA (PARA-) PARAVAX INC.
XX PA (HESK-) HESKA CORP.
XX PI Arfsten A, Dale B, Frank GR, Grieve RB, Heath A;
XX PI Hunter SW, Rushlow KE, Stiegler GL, Yamanaka M;
XX DR WPI; 1996-221762/22.
XX DR N-PSDB; AAT40841.
XX PT DNA encoding flea serine protease and aminopeptidase - useful in
XX PT vaccines to protect animals from flea infestation.
XX PS Claim 71; Page 168; 241pp; English.
XX CC This sequence, PfSP17, represents an N-terminal fragment of serine
XX CC protease PfSP4, isolated from a flea cDNA library by homology with
XX CC conserved serine protease sequences. Other N-terminal sequences
XX CC are given in AAW01187, WO1178 and WO1182. The sequence forms part of
XX CC full-length sequence PfSP4-223 (AAW01165), which also contains the
XX CC sequence PfSP4-52 (AAW01169). The peptide may be used in a vaccine
XX CC for protection of domestic animals from flea infestation, or in
XX CC isolation of protease-inhibitors, which may be used in controlled
XX CC release formulations to reduce the flea burden on and around the
XX CC animal. The inhibitors may be included in insecticidal compositions
XX CC to increase efficacy of other active agents, by reducing proteolytic
XX CC activity in the flea midgut.
XX SQ Sequence 64 AA;
    Query Match 80.9%; Score 38; DB 17; Length 64;
    Best Local Similarity 75.0%; Pred. No. 10;
    Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
DB 35 ILNERWVL 42
RESULT 12
AAB50593
XX ID AAB50593 standard; Protein: 64 AA.
XX AC AAB50593;
XX DT 19-MAR-2001 (first entry)
XX DE Flea serine protease nFSP17 N-terminal protein sequence #75.
XX KW Flea; cat flea; serine protease; aminopeptidase; cysteine protease;
XX KW flea infestation; proteolytic; insecticide; vaccine; cat; dog; flea;
XX KW Ctenocephalides felis; Ctenocephalides canis; Pulex irritans;
XX KW ectoparasite.
XX OS Ctenocephalides felis.
XX PN US6150125-A.
XX PD 21-NOV-2000.
XX PF 24-APR-1996; 96US-0639075.
XX PR 13-DEC-1991; 91US-0806482.
XX PR 18-OCT-1994; 94US-0326773.
XX PR 07-JUN-1995; 95US-0482130.
XX PR 07-JUN-1995; 95US-0484211.
XX PR 07-JUN-1995; 95US-0485443.
XX PR 07-JUN-1995; 95US-0485455.
XX PR 15-AUG-1997; 97WO-US14442.
XX PA (HESK-) HESKA CORP.
XX PI Gaines PJ, Silver G, Rushlow KE, Hunter SW, Frank GR, Stiegler GL;
XX PI Grieve RB;
XX XX WPI; 2001-136374/14.
XX DR N-PSDB; AAC90849.
XX PT New isolated flea proteins with proteolytic activity, useful for
XX PT preventing and reducing flea infestations in mammals especially cats
XX PT and dogs -
XX PS Claim 1; Column 51-52; 150pp; English.
XX CC The present invention describes isolated flea serine protease,
XX CC aminopeptidase and cysteine protease proteins (I). Also described is a
XX CC method for identifying a compound (II) capable of inhibiting flea
XX CC protease activity comprising: (a) contacting (I) with a protease
XX CC substrate and a putative inhibitory compound, where (I) has proteolytic
XX CC activity in the absence of the compound; and (b) determining if the
XX CC compound inhibits protease activity by detecting cleavage of the
XX CC protease substrates; where decreased cleavage of the protease substrate
XX CC indicates an inhibitory compound. (I), nucleic acid molecules encoding
XX CC (I), and antibodies immunospecific for (I) and (II) are useful for
XX CC preventing and reducing flea infestations, particularly the species
XX CC Ctenocephalides felis (cat fleas), Ctenocephalides canis (dog fleas) and
XX CC Pulex irritans, in animals, preferably cats and dogs. They are also
XX CC useful for reducing infestation by other ectoparasites, preferably
XX CC mosquitoes, midges, sandflies, blackflies, ticks and Rhodnius. AAC90818
XX CC to AAC90913 and AAB50551 to AAB50644 represent sequences used in the
XX CC exemplification of the present invention.
XX SQ Sequence 64 AA;
    Query Match 80.9%; Score 38; DB 22; Length 64;
    Best Local Similarity 75.0%; Pred. No. 10;
    Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
DB 35 ILNERWVL 42
RESULT 13
AAW01165
XX ID AAW01165 standard; Protein: 223 AA.
XX AC AAW01165;
XX DT 12-MAR-1997 (first entry)
XX DE Serine protease PfSP4-223.
XX KW Flea; midgut; serine protease; PfSP4-223; vaccine;
XX KW domestic animal; infestation; insecticide; protease-inhibitor;
XX KW controlled release formulation; synergist.
XX OS Siphonaptera sp.
XX PN Key Location/Qualifiers
XX FT Peptide 1..4 /note= "Signal peptide"
XX FT Peptide 1..64 /note= "Sequence AAW01196 (PfSP17, claim 71)"
XX FT Peptide 4..28 /note= "Similar to part of sequence AAW01187 (claim 71)"
XX FT Peptide 5..8 /note= "Conserved serine protease motif"
XX FT Peptide 5..28 /note= "Sequence AAW01182 (claim 71)"

```

FT Peptide 5..29 /note= "Sequence AAW01178 (claim 71)"
 FT Peptide 41..47 /note= "Conserved region around active site His"
 FT Active-site 46 /note= "Active site His residue"
 FT Misc-difference 90 /note= "Conserved Asp residue"
 FT Peptide 124..126 /note= "Conserved serine protease motif"
 FT Peptide 124..175 /note= "Sequence AAW01169 (claim 71)"
 FT Misc-difference 152 /note= "Conserved Cys residue"
 FT Misc-difference 165 /note= "Conserved Cys residue"
 FT Peptide 174..182 /note= "Conserved region around active site Ser"
 FT Active-site 179 /note= "Active site Ser residue"
 FT Peptide 175..182 /note= "Conserved sequence (AAW01226)"
 XX
 PN W09611706-A1.
 XX
 XX 25-APR-1996.
 XX
 XX 18-OCT-1995; 95WO-US14442.
 XX
 PR 07-JUN-1995; 95US-0485455.
 PR 18-OCT-1994; 94US-0326773.
 PR 07-JUN-1995; 95US-0482130.
 PR 07-JUN-1995; 95US-0484211.
 PR 07-JUN-1995; 95US-0485443.
 XX
 PA (PARA-) PARAVAX INC.
 PA (HESK-) HESKA CORP.
 XX
 PI Arfsten A, Dale B, Frank GR, Grieve RB, Heath A;
 PI Hunter SW, Rushlow KE, Stiegler GL, Yamanaka M;
 XX
 DR WPI; 1996-221762/22.
 DR N-PSDB; AAT40818.
 XX
 PT DNA encoding Flea serine protease and aminopeptidase - useful in
 PT vaccines to protect animals from flea infestation.
 XX
 XX Claim 71; Page 137; 24lpp; English.
 PS
 CC This sequence (PfSP4-223) represents a flea midgut serine protease
 CC encoded by the nSP4-672 gene, isolated from a flea cDNA library by
 CC homology with conserved serine protease sequences (e.g. AAW01226).
 CC Although the entire sequence is not highly conserved to that of
 CC known serine proteases, it contains a conserved IVGG sequence,
 CC active site His and Ser residues, a conserved GWG sequence, and
 CC conserved Asp and Cys residues. The sequence contains sequence
 CC AAW01169 (PfSP4-52), which spans 2 conserved serine protease
 CC sequences, N-terminal sequences AAW01196 (PfSP17), AAW01178 and
 CC AAW01182, and a sequence similar to part of AAW01187. The protein shows
 CC homology to an Anopheles gambiae chymotrypsin-1 precursor. The
 CC protein may be used in a vaccine for protection of domestic animals
 CC from flea infestation, or in isolation of protease-inhibitors, which
 CC may be used in controlled release formulations to reduce the flea
 CC burden on and around the animal. The inhibitors may be included in
 CC insecticidal compositions to increase efficacy of other active
 CC agents, by reducing proteolytic activity in the flea midgut.
 XX
 XX Sequence 223 AA;
 SQ
 Query Match 80.9%; Score 38; DB 17; Length 223;
 Best Local Similarity 75.0%; Pred. No. 41;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
 Db 35 ILNERWVL 42
 ::|||
 35 ILNERWVL 42
 RESULT 14
 AAB50561
 ID AAB50561 standard; Protein; 224 AA.
 XX
 AC AAB50561;
 XX
 DT 19-MAR-2001 (first entry)
 XX
 XX Flea serine protease nSP4-672 protein sequence #17.
 XX
 KW Flea; cat flea; serine protease; aminopeptidase; cysteine protease;
 KW flea infestation; proteolytic; insecticide; vaccine; cat; dog; flea;
 KW Ctenocephalides felis; Ctenocephalides canis; Pulex irritans;
 KW ectoparasite.
 XX
 OS Ctenocephalides felis.
 XX
 XX US6150125-A.
 PN
 XX 21-NOV-2000.
 PD
 XX
 XX 24-APR-1996; 96US-0639075.
 PF
 XX 13-DEC-1991; 91US-0806482.
 PR 18-OCT-1994; 94US-0326773.
 PR 07-JUN-1995; 95US-0482130.
 PR 07-JUN-1995; 95US-0484211.
 PR 07-JUN-1995; 95US-0485443.
 PR 07-JUN-1995; 95US-0485455.
 PR 15-AUG-1997; 97WO-US14442.
 XX
 XX (HESK-) HESKA CORP.
 XX
 PI Gaines PJ, Silver G, Rushlow KE, Hunter SW, Frank GR, Stiegler GL;
 PI Grieve RB;
 XX
 DR WPI; 2001-136374/14.
 DR N-PSDB; AAC90823.
 XX
 PT New isolated flea proteins with proteolytic activity, useful for
 PT preventing and reducing flea infestations in mammals especially cats
 PT and dogs -
 XX
 PS Claim 1; Column 87-88; 150pp; English.
 CC
 CC The present invention describes isolated flea serine protease,
 CC aminopeptidase and cysteine protease proteins (I). Also described is a
 CC method for identifying a compound (II) capable of inhibiting flea
 CC protease activity comprising: (a) contacting (I) with a protease
 CC substrate and a putative inhibitory compound, where (I) has proteolytic
 CC activity in the absence of the compound; and (b) determining if the
 CC compound inhibits protease activity by detecting cleavage of the
 CC protease substrates; where decreased cleavage of the protease substrate
 CC indicates an inhibitory compound. (I), nucleic acid molecules encoding
 CC (I), and antibodies immunospecific for (I) and (II) are useful for
 CC preventing and reducing flea infestations, particularly the species
 CC Ctenocephalides felis (cat fleas), Ctenocephalides canis (dog fleas) and
 CC Pulex irritans, in animals, preferably cats and dogs. They are also
 CC useful for reducing infestation by other ectoparasites, preferably
 CC mosquitoes, mites, sandflies, blackflies, ticks and Rhodnius. AAC90818
 CC to AAC90913 and AAB50551 to AAB50644 represent sequences used in the
 CC exemplification of the present invention.
 XX
 XX Sequence 224 AA;
 SQ
 Query Match 80.9%; Score 38; DB 22; Length 224;
 Best Local Similarity 75.0%; Pred. No. 41;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

QY      2 LVNERWVL 9
Db      35 ILNERWVL 42

RESULT 15
AAW64261
ID      AAW64261 standard; Protein; 232 AA.
AC      AAW64261;
XX
XX      24-NOV-1998 (first entry)
XX
XX      Kallikrein substrate binding site.
XX
XX      MCP-7; mast cell protease 7; tryptase-7; serine protease; human;
KW      kallikrein; blood clot; anticoagulant; myocardial infarction;
KW      reocclusion; thromboembolism; cerebral embolism; thrombosis;
KW      therapy.
XX
XX      Class - Mammalia.
XX
XX      Key
Binding-site      Location/Qualifiers
FT      141..152
FT      /note= "putative substrate binding site"
XX
XX      WO9824886-A1.
XX
XX      11-JUN-1998.
XX
XX      25-NOV-1997; 97WO-US21620.
XX
XX      04-DEC-1996; 96US-0032354.
XX
XX      (BGHM ) BRIGHAM & WOMENS HOSPITAL.
XX
XX      Stevens RL;
XX
XX      WPI; 1998-333308/29.
XX
XX      New compositions containing tryptase-7, e.g. mouse mast cell
PT      protease-7 - are used to treat clot formation in e.g. myocardial
PT      infarction, reocclusion following angioplasty or pulmonary
PT      thromboembolism
XX
XX      Example; Page 78; 92pp; English.
XX
XX      This polypeptide comprises the loop regions in the vicinity of the
CC      putative substrate binding site of kallikrein. The crystallographic
CC      structure of this region was used as a template structure to model
CC      the structure of the substrate-binding pocket of mouse mast cell
CC      protease 7 (mMCP-7, see AAW64233). The invention relates to mMCP-7
CC      and related tryptase-7 proteases that can be used to prevent or
CC      inhibit fibrin clot formation. Such proteases can be used to treat
CC      disorders mediated by undesirable thrombus clot formation such as
CC      myocardial infarction and reocclusion following angioplasty of
CC      blood clots associated with pulmonary thromboembolism, deep vein
CC      thrombosis, cerebral embolism, renal vein and peripheral arterial
CC      thrombosis.
XX
XX      Query Match      80.9%; Score 38; DB 19; Length 232;
Best Local Similarity 77.8%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 VLVNERWVL 9
Db      29 VLVNPKWVL 37

RESULT 16
AAW10694
ID      AAW10694 standard; Protein; 260 AA.
AC      AAW10694;
XX
XX      08-APR-1997 (first entry)
XX
XX      Human recombinant neuropsin, used for antibody production.
DE
XX      Antibody; Alzheimer's disease; neurodegenerative; epitope; epilepsy;
KW      recombinant.
XX
XX      Homo sapiens.
XX
XX      Key
Location/Qualifiers
FT      95..215
FT      /note= "claimed fragment of the protein, used for
FT      antibody production"
XX
XX      JP08245700-A.
XX
XX      24-SEP-1996.
XX
XX      14-MAR-1995; 95JP-0083154.
XX
XX      14-MAR-1995; 95JP-0083154.
XX
XX      (IGAK-) IGAKU SEIBUTSUGAKU KENKYUSHO KK.
XX
XX      WPI; 1996-482259/48.
DR      N-PSDB; AAT48519.
XX
XX      Antibody against neuropsin - used in the treatment of Alzheimer's
PT      disease and epilepsy
XX
XX      Claim 1; Page 7; 9pp; Japanese.
XX
XX      AAW10694 is a recombinantly produced human neuropsin protein. The
CC      main invention relates to an antibody against the neuropsin protein,
CC      in particular a claimed fragment of the neuropsin protein (see AAW10695
CC      and features table). The antibody can be used to treat
CC      neurodegenerative disorders such as Alzheimer's disease and epilepsy.
XX
XX      Query Match      80.9%; Score 38; DB 17; Length 260;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 VLVNERWVL 9
Db      61 VLVGDRWVL 69

RESULT 17
AAW12393
ID      AAW12393 standard; Protein; 260 AA.
XX
XX      AAW12393;
AC
XX
XX      15-MAY-1997 (first entry)
XX
XX      Mouse neuropsin protein.
DE
XX
XX      Mouse; neuropsin; hippocampus; lambda gtl0; primer; PCR; amplification;
KW      polymerase chain reaction; serine protease domain; nerve growth factor;
KW      NGF; insect cell; virus; expression vector; transfection;
KW      cerebral disease.
XX
XX      Mus musculus.
XX
XX      JP08311099-A.
XX

```

PD 26-NOV-1996.
 XX
 PF 13-MAR-1996; 96JP-0056367.
 XX
 PR 14-MAR-1995; 95JP-0054584.
 XX
 XX (SHIO/) SHIOZAKA S.
 XX
 XX WPI: 1997-061812/06.
 DR N-PSDB; AAT63251.
 XX
 PT Nucleic acid encoding neuropsin - for producing neuropsin, useful
 for diagnosis and treatment of cerebral disease
 XX
 XX Claim 1; Page 6-7; 9pp; Japanese.
 XX
 CC This is the amino acid sequence of a novel mouse protein designated
 CC neuropsin. The encoding gene was isolated from a mouse hippocampal
 CC cDNA library in lambda gt10 using a cloned, amplified fragment of
 CC the gene (clone B41; AAT63254). This fragment was amplified using
 CC primers AAT63252-3. The primers were synthesised based on the serine
 CC protease domain of nerve growth factor (NGF)-gamma. The screen isolated
 CC 6 positive clones, of which clone NP5 contained the longest insert
 CC (this sequence). The protein has a molecular weight of around 26 kD.
 CC It has 43% homology with EGF-BP, 41% with NGF-gamma; 39% with
 CC NGF-alpha; 38% with trypsin and 18% with tPA. The protein can be used
 CC for clinical diagnosis and treatment of cerebral diseases.
 XX
 SQ Sequence 260 AA;
 Query Match 80.9%; Score 38; DB 18; Length 260;
 Best Local Similarity 77.8%; Pred. No. 48;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 VLVNERWVL 9
 Db 61 VLVGDRWVL 69
 III :IIII
 RESULT 18
 AAW46773
 ID AAW46773 standard; Protein; 233 AA.
 AC
 AC AAW46773;
 XX
 DT 04-JUN-1998 (first entry)
 XX
 DE Amino acid sequence of Salmonase.
 XX
 KW Salmonase: venom gland; Korean viper; Salmosa; serine protease;
 KW Agkistrodon halys brevicaudus; direct-acting; fibrinolytic;
 KW thrombolytic agent; haemostatic agent.
 XX
 OS Agkistrodon halys.
 XX
 FH key Location/Qualifiers
 FT Misc-difference 71 /note- "encoded by CAC"
 FT Misc-difference 176 /note- "encoded by AAC"
 FT
 FT
 XX EP814164-A2.
 XX
 PD 29-DEC-1997.
 XX
 XX 31-OCT-1996; 96EP-0307917.
 PF
 XX 21-JUN-1996; 96KR-0023013.
 PR
 XX (MOGA-) MOGAM BIOTECHNOLOGY RES INST.
 PA
 XX Chung K, Hwang J, Kim D, Koh Y, Moon H, Yun Y;
 PI
 XX

DR WPI: 1998-044340/05.
 DR N-PSDB; AAV16367.
 XX
 PT Salmonase, a direct-acting fibrinolytic serine protease - encoded by
 PT cDNA from snake venom gland library
 XX
 XX Claim 3; Page 25; 28pp; English.
 PS
 XX The present sequence represents a novel direct-acting fibrinolytic
 CC serine protease, designated Salmonase. The DNA sequence was isolated
 CC from a cDNA library constructed from nucleic acid isolated from the
 CC venom gland of a Korean viper Salmosa (Agkistrodon halys brevicaudus).
 CC Salmonase protein comprises two subunits, and has a molecular weight
 CC of about 26 kD and an isoelectric point of 5.04. Salmonase is useful
 CC as a thrombolytic or haemostatic agent.
 XX
 SQ Sequence 233 AA;
 Query Match 78.7%; Score 37; DB 19; Length 233;
 Best Local Similarity 75.0%; Pred. No. 64;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LVNERWVL 9
 Db 29 LINEEWVL 36
 I:III IIII
 RESULT 19
 AAG79000
 ID AAG79000 standard; Protein; 233 AA.
 AC
 AC AAG79000;
 XX
 DT 28-JAN-2002 (first entry)
 XX
 DE Mamushi fibrinolytic enzyme, brevinase.
 XX
 KW Fibrinolytic enzyme; brevinase; thermostable; thrombolytic agent;
 KW mamushi.
 XX
 OS Agkistrodon blomhoffi brevicaudus.
 XX
 PN KR2001045716-A.
 XX
 PD 05-JUN-2001.
 XX
 PF 06-NOV-1999; 99KR-0049115.
 XX
 PR 06-NOV-1999; 99KR-0049115.
 XX
 XX (LEEJ/) LEE J W.
 PA (PARK/) PARK W.
 XX
 XX Lee JW, Park W;
 PI
 XX WPI: 2001-636862/73.
 DR N-PSDB; ABA01202.
 XX
 PT Fibrinolytic enzyme, brevinase, separated from poison of viper,
 PT agkistrodon blomhoffi brevicaudus -
 XX
 XX Claim 2; Page 14-15; 23pp; Korean.
 PS
 XX The present sequence is the protein sequence for fibrinolytic enzyme,
 CC brevinase, which is separated from the poison of Agkistrodon blomhoffi
 CC brevicaudus (mamushi). The enzyme shows stability at high temperatures
 CC and is thus useful in developing thrombolytic agents.
 XX
 SQ Sequence 233 AA;
 Query Match 78.7%; Score 37; DB 22; Length 233;
 Best Local Similarity 75.0%; Pred. No. 64;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I:|I|I|I|
Db 29 LINEWVL 36

RESULT 20
AAW41955
ID AAW41955 standard; Protein; 237 AA.

AC AAW41955;
XX 02-JUL-1998 (first entry)
DE Flea serine protease SEQ ID NO:22.
XX
KW Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;
KW immunoglobulin protease; larvae; host animal.

XX Siphonaptera.
OS
PN WO9740058-A1.
XX 30-OCT-1997.
XX 24-APR-1997; 97WO-US06121.
XX 04-APR-1997; 97US-0042945.
XX 24-APR-1996; 96US-0639075.
XX 15-NOV-1996; 96US-0749699.

XX (HESK-) HESKA CORP.
XX Frank GR, Gaines PJ, Grieve RB, Hunter SW, Rushlow KE;
PI Silver G, Stiegler GL;
XX WPI; 1998-076762/07.
XX N-PSDB; AAV04566.

XX New flea protease genes and proteins - used in vaccine compositions
PT for the prophylaxis and treatment of flea infestation, especially in
PT cats or dogs

XX Claim 2; Page 188-189; 318pp; English.

XX The present sequence represents a novel flea serine protease. The
CC protease, its mimetopes, antibodies (Ab) and inhibitors of the
CC protein, as well as the DNA encoding the protein, may all be used in
CC therapeutic compositions to reduce flea protease activity (especially
CC immunoglobulin protease) and so reduce flea infestation, especially in
CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult
CC fleas which comprises anti-protease antibodies produced by a host animal
CC in response to administration of the protein. Therapeutic compositions
CC may further comprise a compound that reduces haematophagous ectoparasite
CC burden by a method other than by reducing flea immunoglobulin protease
CC activity. The novel flea DNA encoding the protein can also be used to
CC produce recombinant protein, and fragments of it are used as probes and
CC primers for identification and isolation of related sequences, also as
CC antisense, triplex-forming agents and ribozymes for inhibition of the
CC synthesis of the protein. Ab are also useful for screening expression
CC libraries, to purify the protein and to target cytotoxins to fleas.

XX Sequence 237 AA;

Query Match 78.7%; Score 37; DB 19; Length 237;
Best Local Similarity 62.5%; Pred. No. 65;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
I:|I|I|I|
Db 38 IVNDRWIL 45

RESULT 21
ABB60343
ID ABB60343 standard; Protein; 243 AA.

XX ABB60343;
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 7821.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

XX 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

XX Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

XX N-PSDB; ABL04446.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Disclosure; SEQ ID NO 7821; 21pp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

XX The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 243 AA;

Query Match 78.7%; Score 37; DB 22; Length 243;

Best Local Similarity 55.6%; Pred. No. 67;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

I:|I|I|I|

Db 65 VIINEQWIL 73

RESULT 22
AAW41953

ID AAW41953 standard; Protein; 258 AA.

XX AAW41953;

XX 02-JUL-1998 (first entry)

XX Flea serine protease SEQ ID NO:16.

XX Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;
KW immunoglobulin protease; larvae; host animal.

OS Siphonaptera.
 XX WO9740058-A1.
 XX 30-OCT-1997.
 XX 24-APR-1997; 97WO-US06121.
 XX 04-APR-1997; 97US-0042945.
 PR 24-APR-1996; 96US-0639075.
 PR 15-NOV-1996; 96US-0749699.
 XX (HESK-) HESKA CORP.
 PA Frank GR, Gaines PJ, Grieve RB, Hunter SW, Rushlow KE;
 XX Silver G, Stiegler GL;
 PI WPI; 1998-076762/07.
 PT N-PSDB; AAV04562.
 DR
 XX
 XX New flea protease genes and proteins - used in vaccine compositions
 PT for the prophylaxis and treatment of flea infestation, especially in
 PT cats or dogs
 PT
 PS Claim 2; Page 182-183; 318pp; English.
 XX
 CC The present sequence represents a novel flea serine protease. The
 CC protease, its mimetopes, antibodies (Ab) and inhibitors of the
 CC protein, as well as the DNA encoding the protein, may all be used in
 CC therapeutic compositions to reduce flea protease activity (especially
 CC immunoglobulin protease) and so reduce flea infestation, especially in
 CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult
 CC fleas which comprises anti-protease antibodies produced by a host animal
 CC in response to administration of the protein. Therapeutic compositions
 CC may further comprise a compound that reduces haematophagous ectoparasite
 CC activity. The novel flea DNA encoding the protein can also be used to
 CC produce recombinant protein, and fragments of it are used as probes and
 CC primers for identification and isolation of related sequences, also as
 CC antisense, triplex-forming agents and ribozymes for inhibition of the
 CC synthesis of the protein. Ab are also useful for screening expression
 CC libraries, to purify the protein and to target cytotoxins to fleas.
 XX
 SQ Sequence 258 AA;
 Query Match 78.7%; Score 37; DB 19; Length 258;
 Best Local Similarity 62.5%; Pred. No. 71;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LVNERWVL 9
 Db 59 IVNDRWIL 66
 :||:|:|:|
 RESULT 23
 AAW41954
 ID AAW41954 standard; Protein; 258 AA.
 XX
 AC AAW41954;
 DT 02-JUL-1998 (first entry)
 XX
 XX Flea serine protease SEQ ID NO:19.
 XX
 KW Flea; Siphonaptera; protease; vaccine; prophylaxis; infestation;
 KW Immunoglobulin protease; larvae; host animal.
 XX
 OS Siphonaptera.
 XX
 XX WO9740058-A1.
 XX 30-OCT-1997.
 XX

PF 24-APR-1997; 97WO-US06121.
 XX
 PR 04-APR-1997; 97US-0042945.
 PR 24-APR-1996; 96US-0639075.
 PR 15-NOV-1996; 96US-0749699.
 XX (HESK-) HESKA CORP.
 PA Frank GR, Gaines PJ, Grieve RB, Hunter SW, Rushlow KE;
 XX Silver G, Stiegler GL;
 PI WPI; 1998-076762/07.
 PT N-PSDB; AAV04562.
 DR
 XX
 XX New flea protease genes and proteins - used in vaccine compositions
 PT for the prophylaxis and treatment of flea infestation, especially in
 PT cats or dogs
 PT
 PS Claim 2; Page 185-186; 318pp; English.
 XX
 CC The present sequence represents a novel flea serine protease. The
 CC protease, its mimetopes, antibodies (Ab) and inhibitors of the
 CC protein, as well as the DNA encoding the protein, may all be used in
 CC therapeutic compositions to reduce flea protease activity (especially
 CC immunoglobulin protease) and so reduce flea infestation, especially in
 CC cats or dogs. Alternatively, flea larvae may ingest the faeces of adult
 CC fleas which comprises anti-protease antibodies produced by a host animal
 CC in response to administration of the protein. Therapeutic compositions
 CC may further comprise a compound that reduces haematophagous ectoparasite
 CC activity. The novel flea DNA encoding the protein can also be used to
 CC produce recombinant protein, and fragments of it are used as probes and
 CC primers for identification and isolation of related sequences, also as
 CC antisense, triplex-forming agents and ribozymes for inhibition of the
 CC synthesis of the protein. Ab are also useful for screening expression
 CC libraries, to purify the protein and to target cytotoxins to fleas.
 XX
 SQ Sequence 258 AA;
 Query Match 78.7%; Score 37; DB 19; Length 258;
 Best Local Similarity 62.5%; Pred. No. 71;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LVNERWVL 9
 Db 59 IVNDRWIL 66
 :||:|:|:|
 RESULT 24
 AAU03901
 ID AAU03901 standard; Protein; 288 AA.
 XX
 AC AAU03901;
 DT 26-SEP-2001 (first entry)
 XX
 XX Human protease-like polypeptide #3.
 DE
 DE Human; protease; trypsin-like serine protease; enteropeptidase; acrosin;
 KW plasminegen; protein degradation; protein maturation; secretory pathway;
 KW fertility; infectious disease; gene therapy; mental disorder.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FT Misc-difference 67
 FT /note= "Encoded by RCT"
 XX
 PN WO200149864-A1.
 XX
 XX 12-JUL-2001.
 XX
 XX 05-JAN-2001; 2001WO-US00548.
 PF

XX 06-JAN-2000; 2000US-0174686.
XX (LEXI-) LEXICON GENETICS INC.
XX
PI Walke DW, Wilganowski NL, Donoho G, Turner CA;
XX
XX WPI; 2001-432887/46.
XX N-PSDB; AAS06097.
XX
XX Novel polynucleotides encoding human proteins sharing sequence
PT similarity with mammalian proteases, particularly trypsin-like serine
PT proteases, for drug screening, diagnosis and gene therapy of biological
PT disorders.
XX
XX Claim 3; Page 30-31; 34pp; English.
XX
XX The sequence represents a human protein which shares structural
CC similarity with mammalian proteases, particularly trypsin-like serine
CC proteases such as enteropeptidase, plasminogen and acrosin. Proteases
CC cleave protein substrates as part of degradation, maturation and
CC secretory pathways within the body, and are associated with regulating
CC development, modulating cellular processes, fertility and infectious
CC diseases. Hence the new proteins and their associated nucleic acids are
CC useful for drug screening, identification of nucleotide constructs for
CC gene therapy and treatment of mental, biological and medical disorders.
CC The sequences can be used to identify mutations associated with
CC particular diseases, which may subsequently be used in diagnostic and
CC prognostic assays.
XX
XX Query Match 288 AA;
XX
XX Query Match 78.7%; Score 37; DB 22; Length 288;
XX Best Local Similarity 87.5%; Pred. No. 81;
XX Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 51 LVREWRWL 58
II IIIII

RESULT 25
AAU03900
ID AAU03900 standard; Protein: 348 AA.
XX
AC AAU03900;
XX
XX 26-SEP-2001 (first entry)
XX
XX Human protease-like polypeptide #2.
XX
XX Human; protease; trypsin-like serine protease; enteropeptidase; acrosin;
KW plasminogen; protein degradation; protein maturation; secretory pathway;
KW fertility; infectious disease; gene therapy; mental disorder.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 19
FT /note= "Encoded by YAC"
FT Misc-difference 127
FT /note= "Encoded by RCT"
XX
PN WO200149864-A1.
XX
XX 12-JUL-2001.
XX
XX 05-JAN-2001; 2001WO-US00548.
XX
XX 06-JAN-2000; 2000US-0174686.
XX
XX (LEXI-) LEXICON GENETICS INC.
XX

PI Walke DW, Wilganowski NL, Donoho G, Turner CA;
XX
XX WPI; 2001-432887/46.
XX N-PSDB; AAS06096.
XX
XX Novel polynucleotides encoding human proteins sharing sequence
PT similarity with mammalian proteases, particularly trypsin-like serine
PT proteases, for drug screening, diagnosis and gene therapy of biological
PT disorders.
XX
XX Claim 2; Page 29-30; 34pp; English.
XX
XX The sequence represents a human protein which shares structural
CC similarity with mammalian proteases, particularly trypsin-like serine
CC proteases such as enteropeptidase, plasminogen and acrosin. Proteases
CC cleave protein substrates as part of degradation, maturation and
CC secretory pathways within the body, and are associated with regulating
CC development, modulating cellular processes, fertility and infectious
CC diseases. Hence the new proteins and their associated nucleic acids are
CC useful for drug screening, identification of nucleotide constructs for
CC gene therapy and treatment of mental, biological and medical disorders.
CC The sequences can be used to identify mutations associated with
CC particular diseases, which may subsequently be used in diagnostic and
CC prognostic assays.
XX
XX Query Match 348 AA;
XX
XX Query Match 78.7%; Score 37; DB 22; Length 348;
XX Best Local Similarity 87.5%; Pred. No. 1e+02;
XX Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 111 LVREWRWL 118
II IIIII

RESULT 26
AAG31251
ID AAG31251 standard; Protein: 1342 AA.
XX
AC AAG31251;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 37498.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999;	99US-0132048.
PR 30-APR-1999;	99US-0132407.
PR 04-MAY-1999;	99US-0132484.
PR 05-MAY-1999;	99US-0132485.
PR 06-MAY-1999;	99US-0132486.
PR 06-MAY-1999;	99US-0132487.
PR 07-MAY-1999;	99US-0132863.
PR 11-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 14-MAY-1999;	99US-0134370.
PR 18-MAY-1999;	99US-0134768.
PR 19-MAY-1999;	99US-0134941.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143342.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 06-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.

```
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 78.7%; Score 37; DB 21; Length 1342;
Best Local Similarity 66.7%; Pred. No. 4.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
   |||:
Db 1129 VLANKRWVI 1137

RESULT 27
AAG31250
ID AAG31250 standard; Protein: 1382 AA.
AC AAG31250;
XX
XX DT 17-OCT-2000 (first entry)
XX
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 37497.
XX
XX KW Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX OS Arabidopsis thaliana.
XX
XX PN EP1033405-A2.
XX
XX PD 06-SEP-2000.
XX
XX PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135829.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.
XX 18-JUN-1999; 99US-0139463.
XX 18-JUN-1999; 99US-0139750.
XX 18-JUN-1999; 99US-0139763.
XX 21-JUN-1999; 99US-0139817.
XX 22-JUN-1999; 99US-0139899.
XX 23-JUN-1999; 99US-0140353.
XX 23-JUN-1999; 99US-0140354.
XX 28-JUN-1999; 99US-0140695.
XX 28-JUN-1999; 99US-0140823.
XX 30-JUN-1999; 99US-0140991.
XX 30-JUN-1999; 99US-0141287.
XX 01-JUL-1999; 99US-0141842.
XX 01-JUL-1999; 99US-0142154.
XX 02-JUL-1999; 99US-0142055.
XX 06-JUL-1999; 99US-0142390.
XX 08-JUL-1999; 99US-0142803.
XX 09-JUL-1999; 99US-0142920.
XX 12-JUL-1999; 99US-0142977.
XX 13-JUL-1999; 99US-0143542.
XX 14-JUL-1999; 99US-0143624.
XX 15-JUL-1999; 99US-0144005.
XX 16-JUL-1999; 99US-0144085.
XX 16-JUL-1999; 99US-0144086.
XX 19-JUL-1999; 99US-0144325.
XX 19-JUL-1999; 99US-0144331.
XX 19-JUL-1999; 99US-0144332.
XX 19-JUL-1999; 99US-0144333.
XX 19-JUL-1999; 99US-0144334.
XX 19-JUL-1999; 99US-0144335.
XX 20-JUL-1999; 99US-0144352.
XX 20-JUL-1999; 99US-0144632.
XX 20-JUL-1999; 99US-0144884.
XX 21-JUL-1999; 99US-0144814.
XX 21-JUL-1999; 99US-0145086.
XX 21-JUL-1999; 99US-0145088.
XX 22-JUL-1999; 99US-0145085.
XX 22-JUL-1999; 99US-0145087.
XX 22-JUL-1999; 99US-0145089.
XX 22-JUL-1999; 99US-0145192.
XX 23-JUL-1999; 99US-0145145.
```

PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.

PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 78.7%; Score 37; DB 21; Length 1382;
Best Local Similarity 66.7%; Pred. No. 4.6e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
||| :|||:
Db 1169 VLANKRWVI 1177

RESULT 28

AAG31249

ID AAG31249 standard; Protein; 1421 AA.

XX AC AAG31249;

XX DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 37496.

DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX PN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-0301439.

XX PR 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0143055.
PR 06-JUL-1999; 99US-0143390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 20-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 18-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.

PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 78.7%; Score 37; DB 21; Length 1421;
 Best Local Similarity 66.7%; Pred. No. 4.8e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 || |::|||
 DB 1208 VLANKRWVI 1216

RESULT 29
 AAB21302
 ID AAB21302 standard; Protein; 111 AA.

XX AAB21302;
 XX AC
 XX 02-FEB-2001 (first entry)
 XX DE Human KLK-L5 protein #2.

XX KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6;
 KW kallikrein-like protein; serine protease; cytosolic; cancer;
 XX prostrate cancer.

XX OS Homo sapiens.
 XX PN WO200053776-A2.
 XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-CA00258.
 XX PR 11-MAR-1999; 99US-0124260.
 PR 01-APR-1999; 99US-0127386.
 PR 21-JUL-1999; 99US-0144919.
 XX (MOUN) MOUNT SINAI HOSPITAL.

XX Youssef GM, Diamandis EP;
 XX WPI: 2000-587440/55.
 DR N-PSDB; AAA95944.

XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
 PT protein mediated disorders, especially cancer. -

PS Claim 12; Page 172; 184pp; English.

CC The present sequence is one of four alternatively spliced kallikrein-like
 CC proteins encoded by the human KLK-L4 gene. Kallikreins and
 CC kallikrein-like proteins are a subgroup of the serine protease enzyme
 CC family. They catalyze the selective cleavage of specific polypeptide
 CC precursors to release peptides with potent biological activity. Nucleic
 CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
 CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
 CC treatment, monitoring and diagnosis of cancers, especially prostate
 CC cancer. They can also be used to identify a substance that can associate
 CC with or mediate the biological activity of the proteins. Antibodies can
 CC be used to treat conditions mediated by the kallikrein-like proteins.

XX Sequence 111 AA;

Query Match 76.6%; Score 36; DB 21; Length 111;
 Best Local Similarity 66.7%; Pred. No. 42;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 ||:: ||||

DB 50 VLIDHRWVL 58

RESULT 30
 AAY28642
 ID AAY28642 standard; Protein; 162 AA.

XX AC AAY28642;

XX 03-NOV-1999 (first entry)

XX Human secreted protein from cDNA clone HKAFV61.

XX Human secreted protein; human cDNA clone HKAFV61; serine protease family;
 KW Mus musculus neuropsin; serine protease inhibitor; serpin;
 KW IgG domain; IgG-1; IgG-3; albumin; extracellular matrix degradation;
 KW fusion protein; His-tag; cancer; arthritis; cardiovascular disorder;
 KW tissue regeneration; immune system disorder; nervous system disorder.

XX OS Homo sapiens.

XX Key Location/Qualifiers
 FT Peptide 1..17
 FT /label= Signal_peptide
 FT Protein 18..162
 FT /label= Mature_secreted_protein

XX WO9940183-A1.

XX 12-AUG-1999.

XX 04-FEB-1999; 99WO-US02292.

XX 06-FEB-1998; 98US-0073961.

XX (HUMA-) HUMAN GENOME SCI INC.
 XX NI J, Ruben SM;

XX WPI: 1999-508502/42.

XX N-PSDB; AAX80906.

XX New isolated human serine protease and serpin polypeptides, used to
 PT develop products for treating e.g. immune disorders, cancers,
 PT inflammation, transplant rejection or infections, or as food
 PT additives

PS Claim 11; Page 81; 99pp; English.

CC The present sequence is a secreted protein from cDNA clone HKAFV61
 CC which is obtained from human keratinocyte tissue cDNA library. The
 CC protein is a member of serine protease family and shows a high degree
 CC of sequence similarity to Mus musculus neuropsin. The protein, its coding
 CC sequence and its inhibitors are used in the diagnosis and treatment of
 CC disorders related to abnormal level of the protein or mutation in the
 CC nucleotide sequence. The serine protease is used to develop antagonists
 CC and fusion proteins. The fusion of this protein to His-tag, HA-tag, IgG
 CC domains, etc. facilitates purification, fusion to IgG-1, IgG-3 and
 CC albumin increases the half life time in vivo, etc. The inhibitors and
 CC antagonists of serine protease can be used for treating disorders
 CC characterised by degradation of extracellular matrix, e.g. cancer,
 CC arthritis, disorders of cardiovascular system, immune system, nervous
 CC system, etc. The nucleotide and protein sequences can also be used to
 CC differentiate, proliferate and attract cells leading to regeneration of
 CC tissues.

XX Sequence 162 AA;

Query Match 76.6%; Score 36; DB 20; Length 162;
 Best Local Similarity 66.7%; Pred. No. 64;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

```
DE      Novel human diagnostic protein #9688.
XX      Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW      food supplement; medical imaging; diagnostic; genetic disorder.
XX      Homo sapiens.
OS      WO200175067-A2.
XX      11-OCT-2001.
XX      30-MAR-2001; 2001WO-US08631.
XX      31-MAR-2000; 2000US-0540217.
XX      23-AUG-2000; 2000US-0649167.
XX      (HYSE-) HYSEQ INC.
XX      Drmanac RT, Liu C, Tang YT;
XX      WPI: 2001-639362/73.
XX      N-PSDB; AAS73884.
XX      New isolated polynucleotide and encoded polypeptides, useful in
PT      diagnostics, forensics, gene mapping, identification of mutations
PT      responsible for genetic disorders or other traits and to assess
PT      biodiversity.
XX      Claim 20; SEQ ID No 40056; 103pp; English.
XX      The invention relates to isolated polynucleotide (I) and
CC      polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC      polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC      and gene mapping, and in recombinant production of (II). The
CC      polynucleotides are also used in diagnostics as expressed sequence tags
CC      for identifying expressed genes. (I) is useful in gene therapy techniques
CC      to restore normal activity of (II) or to treat disease states involving
CC      (II). (II) is useful for generating antibodies against it, detecting or
CC      quantitating a polypeptide in tissue, as molecular weight markers and as
CC      a food supplement. (II) and its binding partners are useful in medical
CC      imaging of sites expressing (II). (I) and (II) are useful for treating
CC      disorders involving aberrant protein expression or biological activity.
CC      The polypeptide and polynucleotide sequences have applications in
CC      diagnostics, forensics, gene mapping, identification of mutations
CC      responsible for genetic disorders or other traits to assess biodiversity
CC      and to produce other types of data and products dependent on DNA and
CC      amino acid sequences. ABG00010-ABG30377 represent novel human
CC      diagnostic amino acid sequences of the invention.
CC      Note: The sequence data for this patent did not appear in the printed
CC      specification, but was obtained in electronic format directly from WIPO
CC      at ftp.wipo.int/pub/published_pct_sequences.
XX      SQ Sequence 238 AA;

Query Match      76.6%; Score 36; DB 22; Length 238;
Best Local Similarity 77.8%; Pred. No. 98;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
DB 51 VLVAERWLL 59

RESULT 33
AAB21304
ID AAB21304 standard; Protein: 248 AA.
XX
AC AAB21304;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human KLK-L5 protein #4.
XX

Query Match      76.6%; Score 36; DB 21; Length 184;
Best Local Similarity 66.7%; Pred. No. 73;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
DB 37 VLIDHRWVL 45

RESULT 32
ABG09697
ID ABG09697 standard; Protein: 238 AA.
XX
AC ABG09697;
XX
DT 13-FEB-2002 (first entry)
XX
```

DE 50 VLIDHRWVL 58

RESULT 31

AAB21301

ID AAB21301 standard; Protein: 184 AA.

XX

AC AAB21301;

XX

DT 02-FEB-2001 (first entry)

XX

DE Human KLK-L5 protein #1.

XX

Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6;

KW kallikrein-like protein; serine protease; cytostatic; cancer;

KW prostrate cancer.

XX

Homo sapiens.

OS

WO200053776-A2.

PN

14-SEP-2000.

PD

09-MAR-2000; 2000WO-CA00258.

PF

11-MAR-1999; 99US-0124260.

XX

01-APR-1999; 99US-0127386.

PR

21-JUL-1999; 99US-0144919.

PR

(MOUN) MOUNT SINAI HOSPITAL.

XX

Yousef GM, Diamandis EP;

PI

WPI: 2000-587440/55.

DR

N-PSDB; AAS95944.

DR

New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L

PT protein mediated disorders, especially cancer.

PT

Claim 12; Page 172; 184pp; English.

XX

The present sequence is one of four alternatively spliced kallikrein-like

CC proteins encoded by the human KLK-L4 gene. Kallikreins and

CC kallikrein-like proteins are a subgroup of the serine protease enzyme

CC family. They catalyse the selective cleavage of specific polypeptide

CC precursors to release peptides with potent biological activity. Nucleic

CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,

CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the

CC treatment, monitoring and diagnosis of cancers, especially prostate

CC cancer. They can also be used to identify a substance that can associate

CC with or mediate the biological activity of the proteins. Antibodies can

CC be used to treat conditions mediated by the kallikrein-like proteins.

XX

SQ Sequence 184 AA;

Query Match 76.6%; Score 36; DB 21; Length 184;

Best Local Similarity 66.7%; Pred. No. 73;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

DB 37 VLIDHRWVL 45

RESULT 32

ABG09697

ID ABG09697 standard; Protein: 238 AA.

XX

AC ABG09697;

XX

DT 13-FEB-2002 (first entry)

XX

KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6;
 KW kallikrein-like protein; serine protease; cytosstatic; cancer;
 KW prostrate cancer.

OS Homo sapiens.

XX WO200053776-A2.

PN 14-SEP-2000.

XX 09-MAR-2000; 2000WO-CA00258.

XX 11-MAR-1999; 99US-0124260.

PR 01-APR-1999; 99US-0127386.

PR 21-JUL-1999; 99US-0144919.

XX (MOUN) MOUNT SINAI HOSPITAL.

PI Yousef GM, Diamandis EP;

XX WPI: 2000-587440/55.

DR N-PSDB; AAA95944.

XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
 PT protein mediated disorders, especially cancer. -

PS Claim 12; Page 172; 184pp; English.

XX The present sequence is one of four alternatively spliced kallikrein-like
 CC proteins encoded by the human KLK-L4 gene. Kallikreins and
 CC kallikrein-like proteins are a subgroup of the serine protease enzyme
 CC family. They catalyse the selective cleavage of specific polypeptide
 CC precursors to release peptides with potent biological activity. Nucleic
 CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
 CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
 CC treatment, monitoring and diagnosis of cancers, especially prostate
 CC cancer. They can also be used to identify a substance that can associate
 CC with or mediate the biological activity of the proteins. Antibodies can
 CC be used to treat conditions mediated by the kallikrein-like proteins.

XX Sequence 248 AA;

Query Match 76.6%; Score 36; DB 21; Length 248;
 Best Local Similarity 66.7%; Pred. No. 1e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

DB 50 VLVNERWVL 58

RESULT 34

AAB24032

ID AAB24032 standard; Protein; 248 AA.

XX AAB24032;

XX 25-JAN-2001 (first entry)

DE Human PRO1303 protein sequence SEQ ID NO:33.

KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;

KW identification; tumourigenesis; anticancer; detection.

XX Homo sapiens.

XX WO200053750-A1.

XX 14-SEP-2000.

XX 02-DEC-1999; 99WO-US28551.

XX 08-MAR-1999; 99WO-US05028.

PR 01-SEP-1999; 99WO-US20111.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 01-DEC-1999; 99WO-US28634.

XX (GETH) GENENTECH INC.

PI Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;

XX WPI: 2000-594320/56.

DR N-PSDB; AAC58114.

XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit
 PT the growth of tumors in mammals, and to identify inhibitors of PRO
 PT polypeptide activity or expression -

XX Claim 61; Fig 24; 226pp; English.

XX The present invention describes an antibody that binds to a human
 CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
 CC PRO3434; PRO1927; PRO3567; PRO1295; PRO1303; PRO4344; PRO4354;
 CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2038; and PRO2262. (I) has
 CC anticancer activity and can be used to diagnose tumours in mammals, by
 CC detecting complex formation when the antibody is contacted with test
 CC cells. Increased expression of genes encoding (I) can also be detected
 CC to diagnose tumours. Agents which inhibit the activity of (I),
 CC especially the antibodies, or an antisense oligonucleotide which
 CC hybridises to genes encoding (I), can be used to inhibit tumour growth,
 CC preferably by inducing cell death. Methods from the present invention
 CC can be used to identify compounds which inhibit the biological activity
 CC of (I). AAC58109 to AAC58102 represent PCR primers and hybridisation
 CC probes used in examples from the present invention for human PRO
 CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
 CC PRO polynucleotide and protein sequences given in the exemplification of
 CC the present invention.

XX Sequence 248 AA;

Query Match 76.6%; Score 36; DB 21; Length 248;
 Best Local Similarity 66.7%; Pred. No. 1e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9

DB 50 VLVNERWVL 58

RESULT 35

AAB24428

ID AAB24428 standard; Protein; 248 AA.

XX AAB24428;

XX 07-NOV-2000 (first entry)

DE Human PRO1303 protein sequence SEQ ID NO:203.

XX Human; PRO; promotion; inhibition; angiogenesis; cardiovascularisation;
 KW diagnosis; trauma; wound; cancer; atherosclerosis; cardiac hypertrophy;
 KW angiogenic; proliferative; cardiac; cardiovascular; antiatherosclerotic;
 KW cytostatic; gene therapy; vaccine.

XX Homo sapiens.

XX WO200032221-A2.

XX 08-JUN-2000.

XX 30-NOV-1999; 99WO-US28313.

XX 01-DEC-1998; 98WO-US25108.

PR 16-DEC-1998; 98US-0112850.

PR 12-JAN-1999; 99US-0115554.

PR 08-MAR-1999; 99WO-US05028.
 PR 12-MAR-1999; 99US-0123957.
 PR 28-APR-1999; 99US-0131445.
 PR 14-MAY-1999; 99US-0134287.
 PR 02-JUN-1999; 99WO-US12252.
 PR 23-JUN-1999; 99US-0141037.
 PR 26-JUL-1999; 99US-0144758.
 PR 20-AUG-1999; 99US-0145698.
 PR 01-SEP-1999; 99WO-US20111.
 PR 08-SEP-1999; 99WO-US20594.
 PR 13-SEP-1999; 99WO-US20944.
 PR 15-SEP-1999; 99WO-US21090.
 PR 15-SEP-1999; 99WO-US21547.
 PR 05-OCT-1999; 99WO-US23089.
 PR 29-OCT-1999; 99US-0162506.
 XX
 PA (GETH) GENENTECH INC.
 XX
 PI Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Hillan KJ, Goddard A;
 PI Godowski FJ, Gurney AL, Klein RD, Kuo SS, Paoni NF, Smith V;
 PI Watanabe CK, Williams PM, Wood WI;
 XX
 DR WPI: 2000-412154/35.
 DR N-PSDB; AAA77671.
 XX
 PT Nucleic acids encoding PRO polypeptides useful for preventing,
 PT diagnosing and treating disorders in mammals -
 PT angiogenic disorders in mammals -
 PT
 XX
 PS Claim 72; Fig 82; 315pp; English.
 XX
 CC The present invention describes nucleic acids encoding PRO polypeptides
 CC useful for preventing, diagnosing and treating disorders in mammals by
 CC cardiovascular, endothelial or angiogenic disorder in mammals by
 CC modulating cell proliferation, angiogenesis and cardiovascularisation,
 CC and for identifying agonists and antagonists of these processes. The
 CC nucleic acids and the proteins they encode may be used in the
 CC prevention, treatment and diagnosis of diseases associated with
 CC inappropriate PRO expression such as cardiovascular, endothelial or
 CC angiogenic disorders in mammals (e.g. atherosclerosis, cancers and
 CC cardiac hypertrophy). For example, the nucleic acids (NCS) and vectors
 CC containing them and the PRO polypeptide may be used to treat disorders
 CC associated with decreased PRO expression. AAA77510 to AAA77721 and
 CC AAB24388 to AAB24435 represent nucleotide and protein sequences used in
 CC the exemplification of the present invention.
 XX
 SQ Sequence 248 AA;
 Query Match 76.6%; Score 36; DB 21; Length 248;
 Best Local Similarity 66.7%; Pred. No. 1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 VLVNERWVL 9
 II:: IIII
 Db 50 VLIDHRWVL 58
 RESULT 36
 AAY99393
 ID AAY99393 standard; Protein; 248 AA.
 XX
 AC AAY99393;
 XX
 DT 08-AUG-2000 (first entry)
 XX
 DE Human PRO1303 (UNQ669) amino acid sequence SEQ ID NO:194.
 XX
 KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
 KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.
 XX
 OS Homo sapiens.
 XX
 PN WO200012708-A2.

XX 09-MAR-2000.
 PD 01-SEP-1999; 99WO-US20111.
 XX
 PF 01-SEP-1998; 99US-0098716.
 XX 01-SEP-1998; 99US-0098749.
 PR 01-SEP-1998; 99US-0098750.
 PR 02-SEP-1998; 99US-0098803.
 PR 02-SEP-1998; 99US-0098821.
 PR 02-SEP-1998; 99US-0098843.
 PR 09-SEP-1998; 99US-0099536.
 PR 09-SEP-1998; 99US-0099596.
 PR 09-SEP-1998; 99US-0099602.
 PR 09-SEP-1998; 99US-0099642.
 PR 10-SEP-1998; 99US-0099741.
 PR 10-SEP-1998; 99US-0099754.
 PR 10-SEP-1998; 99US-0099763.
 PR 10-SEP-1998; 99US-0099792.
 PR 10-SEP-1998; 99US-0099808.
 PR 10-SEP-1998; 99US-0099812.
 PR 10-SEP-1998; 99US-0099815.
 PR 10-SEP-1998; 99US-0099816.
 PR 15-SEP-1998; 99US-0100385.
 PR 15-SEP-1998; 99US-0100388.
 PR 15-SEP-1998; 99US-0100390.
 PR 16-SEP-1998; 99US-0100584.
 PR 16-SEP-1998; 99US-0100627.
 PR 16-SEP-1998; 99US-0100661.
 PR 16-SEP-1998; 99US-0100662.
 PR 16-SEP-1998; 99US-0100664.
 PR 17-SEP-1998; 99US-0100683.
 PR 17-SEP-1998; 99US-0100684.
 PR 17-SEP-1998; 99US-0100710.
 PR 17-SEP-1998; 99US-0100711.
 PR 17-SEP-1998; 99US-0100919.
 PR 17-SEP-1998; 99US-0100930.
 PR 18-SEP-1998; 99US-0100848.
 PR 18-SEP-1998; 99US-0100849.
 PR 18-SEP-1998; 99US-0101014.
 PR 18-SEP-1998; 99US-0101068.
 PR 18-SEP-1998; 99US-0101071.
 PR 22-SEP-1998; 99US-0101279.
 PR 23-SEP-1998; 99US-0101471.
 PR 23-SEP-1998; 99US-0101472.
 PR 23-SEP-1998; 99US-0101474.
 PR 23-SEP-1998; 99US-0101475.
 PR 23-SEP-1998; 99US-0101476.
 PR 23-SEP-1998; 99US-0101477.
 PR 23-SEP-1998; 99US-0101479.
 PR 24-SEP-1998; 99US-0101738.
 PR 24-SEP-1998; 99US-0101741.
 PR 24-SEP-1998; 99US-0101743.
 PR 24-SEP-1998; 99US-0101915.
 PR 24-SEP-1998; 99US-0101916.
 PR 29-SEP-1998; 99US-0102207.
 PR 29-SEP-1998; 99US-0102240.
 PR 29-SEP-1998; 99US-0102307.
 PR 29-SEP-1998; 99US-0102330.
 PR 29-SEP-1998; 99US-0102331.
 PR 30-SEP-1998; 99US-0102484.
 PR 30-SEP-1998; 99US-0102487.
 PR 30-SEP-1998; 99US-0102570.
 PR 30-SEP-1998; 99US-0102571.
 PR 01-OCT-1998; 99US-0102684.
 PR 01-OCT-1998; 99US-0102687.
 PR 02-OCT-1998; 99US-0102965.
 PR 06-OCT-1998; 99US-0103258.
 PR 06-OCT-1998; 99US-0103449.
 PR 07-OCT-1998; 99US-0103314.
 PR 07-OCT-1998; 99US-0103315.
 PR 07-OCT-1998; 99US-0103328.

Db 50 VLIDHRWL 58
|||||

RESULT 38

AAB66142

ID AAB66142 standard; protein; 248 AA.

XX AC AAB66142;

XX DT 02-APR-2001 (first entry)

XX DE Protein of the invention #54.

XX KW Secreted; transmembrane; gene therapy.

XX OS Unidentified.

XX PN WO200078961-A1.

XX PD 28-DEC-2000.

XX PF 18-FEB-2000; 2000WO-US04342.

XX PR 23-JUN-1999; 99US-0141037.

XX PR 20-JUL-1999; 99US-0144758.

XX PR 26-JUL-1999; 99US-0145698.

XX PR 01-SEP-1999; 99WO-US20111.

XX PR 29-OCT-1999; 99US-0162506.

XX PR 30-NOV-1999; 99WO-US28313.

XX PR 02-DEC-1999; 99WO-US28551.

XX PR 16-DEC-1999; 99WO-US30095.

XX PR 05-JAN-2000; 2000WO-US00219.

XX PR 06-JAN-2000; 2000WO-US00376.

XX PA (GETH) GENENTECH INC.

XX PI Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

XX PI Gao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

XX PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

XX PI Watanabe CK, Williams PM, Wood WI;

XX WPI: 2001-071395/08.

XX Secreted and transmembrane proteins and nucleic acids designated PRO,

XX useful as hybridization probes, in chromosome and gene mapping and gene

XX therapy -

XX Claim 1; Fig 108; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.

XX These proteins and the DNA encoding them may be used as hybridization

XX probes, in chromosome and gene mapping and in the generation of

XX anti-sense RNA and DNA. They may also be used to generate either

XX transgenic animals or knockout animals which are in turn useful for

XX development and screening of therapeutically useful reagents.

XX The nucleic acids may also be used in gene therapy.

XX Sequence 248 AA;

XX Query Match 76.6%; Score 36; DB 22; Length 248;

XX Best Local Similarity 66.7%; Pred. NO. 1e+02;

XX Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

XX QY 1 VLVNERWL 9

XX Db 50 VLIDHRWL 58

XX RESULT 39

AAB21303

ID AAB21303 standard; Protein; 254 AA.

XX

AC AAB21303;

XX DT 02-FEB-2001 (first entry)

XX DE Human KLK-L5 protein #3.

XX KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6;

XX KW kallikrein-like protein; serine protease; cytostatic; cancer;

XX KW prostrate cancer.

XX OS Homo sapiens.

XX PN WO200053776-A2.

XX PD 14-SEP-2000.

XX PF 09-MAR-2000; 2000WO-CA00258.

XX PR 11-MAR-1999; 99US-0124260.

XX PR 01-APR-1999; 99US-0127386.

XX PR 21-JUL-1999; 99US-0144919.

XX PA (MOUN) MOUNT SINAI HOSPITAL.

XX PI Yousef GM, Diamandis EP;

XX WPI: 2000-587440/55.

XX N-PSDB; AAR95944.

XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L

XX protein mediated disorders, especially cancer.

XX Claim 12; Page 172; 184pp; English.

XX The present sequence is one of four alternatively spliced kallikrein-like

XX proteins encoded by the human KLK-L4 gene. Kallikreins and

XX kallikrein-like proteins are a subgroup of the serine protease enzyme

XX family. They catalyze the selective cleavage of specific polypeptide

XX precursors to release peptides with potent biological activity. Nucleic

XX acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,

XX KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the

XX treatment, monitoring and diagnosis of cancers, especially prostate

XX cancer. They can also be used to identify a substance that can associate

XX with or mediate the biological activity of the proteins. Antibodies can

XX be used to treat conditions mediated by the kallikrein-like proteins.

XX Query Match 76.6%; Score 36; DB 21; Length 254;

XX Best Local Similarity 66.7%; Pred. NO. 1.1e+02;

XX Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

XX QY 1 VLVNERWL 9

XX Db 50 VLIDHRWL 58

XX RESULT 40

XX AAR55757

XX ID AAR55757 standard; Protein; 258 AA.

XX AC AAR55757;

XX DT 17-NOV-1994 (first entry)

XX DE Serine protease RNK Met-1.

XX KW Serine protease; RNK Met-1; cytolytic granule; leukemia;

XX KW large granular lymphocyte; RNK-16; Met-ase; tumor; diagnosis;

XX KW DNA probe; hybridization.

XX OS Rattus sp.

XX

FH Key Location/Qualifiers
FT Peptide 1..20
FT /label= Sig_peptide
FT Active-site 61
FT Active-site 107
FT Modified-site 174..176
FT /note= "putative N-glycosylation site"
FT Binding-site 198
FT /note= "determines specificity of substrate binding"
FT 204
FT Active-site 220..222
FT Binding-site /note= "determine specificity of substrate binding"
FT 225..227
FT Modified-site /note= "putative N-glycosylation site"
FT Binding-site 233
FT /note= "determines specificity of substrate binding"
FT FT
FT FT
PN W09412647-A.
XX
XX 09-JUN-1994.
PD
XX 03-DEC-1993; 93WO-US11736.
XX
XX 03-DEC-1992; 92US-0990301.
XX
XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
XX
XX Henderson LE, Powers JC, Sayer T, Smyth MJ, Sowder R;
PI Willtrout TA;
XX
XX WPI; 1994-200275/24.
DR N-PSDB; AAQ66909.
XX
XX New serine protease with Met-ase activity - isolated from
PT lymphocyte granules, also related DNA vectors, transformed cells
PT and antibodies.
XX
XX Disclosure; Page 35-36; 57pp; English.
XX
XX A new 30 kDa serine protease, RNK Met-1, was isolated from the
CC cytolitic granules of rat RNK-16 large granular lymphocyte leukemia
CC cells. cDNA was obtained from a rat RNK-16 lambda-gt11 library and
CC sequenced (AAQ66909). The mature protein (sequence AAR55757) has an N-
CC terminal conserved sequence (AAR55761) found in activated serine
CC proteases. The sequences given in AAR55762 and AAR55763 respectively
CC include the active site His-41 and Ser-184 of RNK Met-1 and are also
CC highly conserved. DNA encoding RNK Met-1 can be used as a probe to
CC detect the enzyme in tumor biopsy samples.
XX
XX Sequence 258 AA;
SQ
Query Match 76.6%; Score 36; DB 15; Length 258;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLVNERWVL 9
DB 49 VLVHQRWVL 57
RESULT 41
ABB63123
ID ABB63123 standard; Protein; 547 AA.
XX
XX ABB63123;
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 16161.
DE
XX

KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX W0200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX Venter JC, Adams M, Li PWD, Myers EW;
PI WPI; 2001-656860/75.
XX N-PSDB; ABL07226.
DR
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
PT
XX Disclosure; SEQ ID NO 16161; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 547 AA;
SQ
Query Match 76.6%; Score 36; DB 22; Length 547;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 VLVNERW 7
DB 405 VLLNERW 411
RESULT 42
ABB60547
ID ABB60547 standard; Protein; 580 AA.
XX
XX ABB60547;
AC
XX 26-MAR-2002 (first entry)
DT
XX Drosophila melanogaster polypeptide SEQ ID NO 8433.
DE
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
KW
XX Drosophila melanogaster.
OS
XX W0200171042-A2.
PN
XX 27-SEP-2001.
PD
XX 23-MAR-2001; 2001WO-US09231.
PF
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX

XX (PEKE) PE CORP NY.
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
DR N-PSDB; ABL04650.
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX Disclosure; SEQ ID NO 8433; 21pp + Sequence Listing; English.
PS
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 580 AA;
SQ
Query Match 76.6%; Score 36; DB 22; Length 580;
Best Local Similarity 55.6%; Pred. No. 2.6e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 VLVNERWVL 9
DB 383 VILNRRWIL 391
I:::I I::I
RESULT 43
AAE08323
ID AAE08323 standard; peptide; 9 AA.
XX
AC AAE08323;
DT
XX 01-NOV-2001 (first entry)
XX Human stratum corneum chymotrypsin enzyme peptide #88 (residues 61-69).
XX Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.
XX Homo sapiens.
XX WO200159158-A1.
PN
XX 16-AUG-2001.
XX
XX 07-FEB-2001; 2001WO-US03977.
XX
XX 11-FEB-2000; 2000US-0502600.
XX (UYAR-) UNIV ARKANSAS.
XX
XX O'brien TJ;
PI
XX WPI; 2001-514676/56.
DR
XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -
PT
XX Disclosure; Page 122; 127pp; English.
PS
XX The invention relates to diagnosing cancer especially ovarian cancer, by

CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.
XX
XX Sequence 9 AA;
SQ
Query Match 74.5%; Score 35; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 4 NERWVL 9
DB 1 NERWVL 6
I I I I I I
RESULT 44
AAAY51621
ID AAY51621 standard; Protein; 60 AA.
XX
AC AAY51621;
XX
XX 02-JUN-2000 (first entry)
DT
XX Rabbit Factor IX protein fragment.
DE
XX RAWIT; RNA amplification with in vitro translation; RAWTS; Factor IX;
KW RNA amplification with transcript sequencing; rabbit.
XX
XX Oryctolagus cuniculus.
OS
XX US6027913-A.
PN
XX 22-FEB-2000.
PD
XX 27-DEC-1994; 94US-0399855.
PF
XX 24-JUL-1989; 89US-0385013.
PR 12-NOV-1993; 93US-0151461.
PR 28-JAN-1988; 88US-0149312.
XX
XX (SOMM/) SOMMER S S.
PA
XX Sommer SS;
PI
XX WPI; 2000-194830/17.
DR N-PSDB; AAZ89011.
DR
XX Producing polypeptides from nucleic acid sequences of interest using a
PT modified RAWIT (RNA amplification with in vitro translation) method -
PT
XX Disclosure; Fig 4; 65pp; English.
PS
XX This invention describes a novel modified RAWIT (RNA amplification with
CC in vitro translation) method, for synthesizing polypeptides encoded by
CC sequences of interest. The method generates a protein from an RNA
CC molecule and comprises (a) incubating the RNA with reverse transcriptase
CC to form cDNA; (b) amplifying the cDNA of step (a) to generate DNA copies
CC of the cDNA in the presence of at least 1 oligonucleotide primer which
CC contains an RNA promoter sequence; (c) transcribing the DNA generated
CC with an RNA polymerase to form RNA transcripts; and (d) translating the
CC RNA transcripts in vitro to generate the protein. This sequence
CC represents a fragment of the rabbit Factor IX protein which is used to
CC illustrate the RAWTS (RNA amplification with transcript sequencing)
CC method of the invention.
XX
XX Sequence 60 AA;
SQ
Query Match 74.5%; Score 35; DB 21; Length 60;

Best Local Similarity 62.5%; Pred. No. 32;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
DB 10 IVNEKWWV 17
:|||||:

RESULT 45
AAY51623
ID AAY51623 standard; Protein; 60 AA.
XX
AC AAY51623;
XX
DT 02-JUN-2000 (first entry)
XX
DE Bovine Factor IX protein fragment.
XX
KW RAWT; RNA amplification with in vitro translation; RAWTS; Factor IX;
KW RNA amplification with transcript sequencing; bovine.
XX
OS Bos taurus.
XX
PN US6027913-A.
XX
PD 22-FEB-2000.
XX
PF 27-DEC-1994; 9AUS-0399855.
XX
PR 24-JUL-1989; 89US-0385013.
PR 12-NOV-1993; 93US-0151461.
PR 28-JAN-1988; 88US-0149312.
XX
PA (SOMM/) SOMMER S S.
XX
PI Sommer SS;
XX
DR WPI; 2000-194830/17.
XX
PT Producing polypeptides from nucleic acid sequences of interest using a
PT modified RAWIT (RNA amplification with in vitro translation) method -
XX
PS Disclosure; Fig 4; 65pp; English.
XX
CC This invention describes a novel modified RAWIT (RNA amplification with
CC in vitro translation) method, for synthesizing polypeptides encoded by
CC sequences of interest. The method generates a protein from an RNA
CC molecule and comprises (a) incubating the RNA with reverse transcriptase
CC to form cDNA; (b) amplifying the cDNA of step (a) to generate DNA copies
CC of the cDNA in the presence of at least 1 oligonucleotide primer which
CC contains an RNA promoter sequence; (c) transcribing the DNA generated
CC with an RNA polymerase to form RNA transcripts; and (d) translating the
CC RNA transcripts in vitro to generate the protein. This sequence
CC represents a fragment of the bovine Factor IX protein which is used to
CC illustrate the RAWTS (RNA amplification with transcript sequencing)
CC method of the invention.
XX
SQ Sequence 60 AA;

Query Match 74.5%; Score 35; DB 21; Length 60;
Best Local Similarity 62.5%; Pred. No. 32;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
DB 10 IVNEKWWV 17
:|||||:

Search completed: November 6, 2002, 12:05:08
Job time : 25.7778 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 8.11111 Seconds
(without alignments)
27.102 Million cell updates/sec

Title: US-09-905-083-34
Perfect score: 47
Sequence: 1 VLVNERRWVL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA: *
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep: *
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep: *
3: /cgn2_6/ptodata/1/1aa/5A-COMB.pep: *
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep: *
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep: *
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep: *

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Query Score	% Match	Length	ID	Description
1	47	100.0	9	4 US-09-502-600-34	Sequence 34, Appl
2	47	100.0	154	4 US-09-261-416-7	Sequence 7, Appl
3	47	100.0	224	4 US-08-944-483-33	Sequence 33, Appl
4	47	100.0	225	2 US-08-557-146-12	Sequence 12, Appl
5	47	100.0	225	2 US-09-027-337-4	Sequence 4, Appl
6	47	100.0	225	2 US-09-154-344-12	Sequence 12, Appl
7	47	100.0	253	2 US-08-557-146-2	Sequence 2, Appl
8	47	100.0	253	2 US-08-824-874-3	Sequence 3, Appl
9	47	100.0	253	2 US-09-154-344-2	Sequence 2, Appl
10	47	100.0	253	3 US-08-930-188-2	Sequence 2, Appl
11	47	100.0	253	4 US-09-210-084-3	Sequence 3, Appl
12	47	100.0	253	5 PCT-US96-04294-2	Sequence 2, Appl
13	43	91.5	9	4 US-09-502-600-40	Sequence 40, Appl
14	43	91.5	9	4 US-09-502-600-50	Sequence 50, Appl
15	39	83.0	228	4 US-08-944-483-44	Sequence 44, Appl
16	39	83.0	253	6 5223425-8	Patent No. 5223425
17	38	80.9	64	1 US-08-485-455D-75	Sequence 75, Appl
18	38	80.9	64	2 US-08-482-130C-75	Sequence 75, Appl
19	38	80.9	64	2 US-08-484-211C-75	Sequence 75, Appl
20	38	80.9	64	3 US-08-906-769-75	Sequence 75, Appl
21	38	80.9	64	3 US-08-906-616-75	Sequence 75, Appl
22	38	80.9	64	4 US-08-817-795-75	Sequence 75, Appl
23	38	80.9	64	4 US-08-485-443B-75	Sequence 75, Appl
24	38	80.9	64	4 US-08-639-075A-75	Sequence 75, Appl
25	38	80.9	64	4 US-09-012-431-75	Sequence 75, Appl
26	38	80.9	64	4 US-09-012-692-75	Sequence 75, Appl
27	38	80.9	64	4 US-08-906-613-75	Sequence 75, Appl

28	38	80.9	64	5 PCT-US95-14442A-75	Sequence 75, Appl
29	38	80.9	223	1 US-08-485-455D-17	Sequence 17, Appl
30	38	80.9	223	2 US-08-482-130C-17	Sequence 17, Appl
31	38	80.9	223	2 US-08-484-211C-17	Sequence 17, Appl
32	38	80.9	223	4 US-08-817-795-17	Sequence 17, Appl
33	38	80.9	223	4 US-08-485-443B-17	Sequence 17, Appl
34	38	80.9	223	5 PCT-US95-14442A-17	Sequence 17, Appl
35	38	80.9	224	3 US-08-906-769-17	Sequence 17, Appl
36	38	80.9	224	3 US-08-906-616-17	Sequence 17, Appl
37	38	80.9	224	4 US-08-639-075A-17	Sequence 17, Appl
38	38	80.9	224	4 US-09-012-431-17	Sequence 17, Appl
39	38	80.9	224	4 US-09-012-692-17	Sequence 17, Appl
40	38	80.9	224	4 US-08-906-613-17	Sequence 17, Appl
41	38	80.9	232	1 US-08-278-091-8	Sequence 8, Appl
42	38	80.9	232	1 US-08-483-859-8	Sequence 8, Appl
43	38	80.9	232	1 US-08-472-173-8	Sequence 8, Appl
44	38	80.9	232	2 US-08-487-167-8	Sequence 8, Appl
45	38	80.9	232	2 US-08-482-816-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-09-502-600-34
; Sequence 34, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 34
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 58-66 of the SCCE protein
US-09-502-600-34

Query Match 100.0%; Score 47; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 VLVNERRWVL 9
Db 1 VLVNERRWVL 9

RESULT 2

US-09-261-416-7
; Sequence 7, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TAGD-12: A No. 6291663el Transmembrane Serine Protease
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261,416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 7
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of stratum corneum

; OTHER INFORMATION: chymotryptic enzyme (Scce) homologous to similar
; OTHER INFORMATION: domain in TADG-12
US-09-261-416-7

Query Match 100.0%; Score 47; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.15;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 |||||
DB 4 VLVNERWVL 12

RESULT 3

US-08-944-483-33
; Sequence 33, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; TITLE OF INVENTION: OF THE PROSTATE
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-33

Query Match 100.0%; Score 47; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 |||||
DB 29 VLVNERWVL 37

RESULT 4

US-08-557-146-12
; Sequence 12, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
US-08-557-146-12

Query Match 100.0%; Score 47; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
 |||||
DB 30 VLVNERWVL 38

RESULT 5

US-09-027-337-4
; Sequence 4, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirotooshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of Scce homologous to
; OTHER INFORMATION: similar domain in TADG-15

US-09-027-337-4

Query Match 100.0%; Score 47; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
| | | | | | | | | |
DB 30 VLVNERWVL 38

RESULT 6

US-09-154-344-12
; Sequence 12, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
US-09-154-344-12

Query Match 100.0%; Score 47; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.23;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
| | | | | | | | | |
DB 30 VLVNERWVL 38

RESULT 7

US-08-557-146-2
; Sequence 2, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart

; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-557-146-2

Query Match 100.0%; Score 47; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
| | | | | | | | | |
DB 58 VLVNERWVL 66

RESULT 8

US-08-824-874-3
; Sequence 3, Application US/08824874
; Patent No. 5962300
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE: Filed Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0252 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 532504
US-08-824-874-3

Query Match 100.0%; Score 47; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
IIIIIIIIII
DB 58 VLVNERWVL 66

RESULT 9
US-09-154-344-2
Sequence 2, Application US/09154344
Patent No. 5981256
GENERAL INFORMATION:
APPLICANT: Egelrud, Torbjorn
APPLICANT: Hansson, Lennart
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
ENZYME (SCCE)
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case, Patent Department
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2787

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,344
FILING DATE: 16-SEP-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/557,146
FILING DATE: 14-DEC-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-181
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-154-344-2

Query Match 100.0%; Score 47; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
IIIIIIIIII
DB 58 VLVNERWVL 66

RESULT 10
US-08-930-188-2
Sequence 2, Application US/08930188
Patent No. 6093397
GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Little, Sheila P.
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
RELATED NUCLEIC ACIDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/930,188
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Blalock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-930-188-2

Query Match 100.0%; Score 47; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
IIIIIIIIII
DB 58 VLVNERWVL 66

RESULT 11
US-09-210-084-3
Sequence 3, Application US/09210084
Patent No. 6197511
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL KALLIKREIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive

```
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: PastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,084
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 532504
; US-09-210-084-3

Query Match 100.0%; Score 47; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
Db 58 VLVNERWVL 66

RESULT 12
PCT-US96-04294-2
; SEQUENCE 2, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
```

```
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US96-04294-2

Query Match 100.0%; Score 47; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.26;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
Db 58 VLVNERWVL 66

RESULT 13
US-09-502-600-40
; SEQUENCE 40, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; TITLE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US/09/502,600A
; PRIOR FILING DATE: 09/039,211
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 40
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 57-65 of the SCCE protein
; US-09-502-600-40

Query Match 91.5%; Score 43; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 VLVNERWV 8
Db 2 VLVNERWV 9

RESULT 14
US-09-502-600-50
; SEQUENCE 50, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; TITLE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US/09/502,600A
; PRIOR FILING DATE: 09/039,211
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 50
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
```

;
; FEATURE:
; OTHER INFORMATION: Residues 59-67 of the SCCE protein
US-09-502-600-50

Query Match 91.5%; Score 43; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 1 LVNERWVL 8

RESULT 15
US-08-944-483-44
; Sequence 44, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; TITLE OF INVENTION: OF THE PROSTATE
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 228 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-44

Query Match 83.0%; Score 39; DB 4; Length 228;
Best Local Similarity 88.9%; Pred. No. 6.1;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LVNERWVL 9
Db 29 VLVAERWVL 37

RESULT 16
5223425-8
; Patent No. 5223425
; APPLICANT: FLIER, JEFFREY S.; SPIEGELMAN, BRUCE M.; ROSEN,
; BARRY M.; WHITE, TYLER R.
; TITLE OF INVENTION: DNA ENCODING HUMAN ADIPSIN WITH COMPLEMENT
; D ACTIVITY
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/277,963
; FILING DATE: 30-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 34,203
; FILING DATE: 02-APR-1987
; SEQ ID NO: 8
; LENGTH: 253
5223425-8

Query Match 83.0%; Score 39; DB 6; Length 253;
Best Local Similarity 88.9%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
Db 54 VLVAERWVL 62

RESULT 17
US-08-485-455D-75
; Sequence 75, Application US/08485455D
; Patent No. 5712143
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary L.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,455D
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C1-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-485-455D-75

Query Match 80.9%; Score 38; DB 1; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 18
US-08-482-130C-75
; Sequence 75, Application US/08482130C
; Patent No. 5962257
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary L.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/482,130C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C1-1
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-482-130C-75

Query Match 80.9%; Score 38; DB 2; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 19
US-08-484-211C-75
; Sequence 75, Application US/08484211C
; Patent No. 5972645
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary L.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID

; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,211C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-484-211C-75

Query Match 80.9%; Score 38; DB 2; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 20
US-08-906-769-75
; Sequence 75, Application US/08906769
; Patent No. 6077687
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,769
; FILING DATE:
; CLASSIFICATION:

;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/639,075
;; FILING DATE: 24-APR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Connell, Gary J.
;; REGISTRATION NUMBER: 32,020
;; REFERENCE/DOCKET NUMBER: 2618-25-C2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (303) 863-9700
;; TELEFAX: (303) 863-0223
;; INFORMATION FOR SEQ ID NO: 75:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 64 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-906-769-75

Query Match 80.9%; Score 38; DB 3; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 21
US-08-906-616-75
; Sequence 75, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grievé, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,616
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-906-616-75

Query Match 80.9%; Score 38; DB 3; Length 64;

Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 22
US-08-817-795-75
; Sequence 75, Application US/08817795
; Patent No. 6139840
; GENERAL INFORMATION:
; APPLICANT: Grievé, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Frank, Glenn R.
; APPLICANT: Heath, Andrew W.
; APPLICANT: Yamaka, Miles Yamanaka
; APPLICANT: Arfsten, Ann
; APPLICANT: Dale, Beverly
; APPLICANT: Stiegler, Gary
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,795
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14442
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gary J. Connell
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-817-795-75

Query Match 80.9%; Score 38; DB 4; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 23
US-08-485-443B-75
; Sequence 75, Application US/08485443B

Patent No. 6146870
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Hunter, Shirley Wu
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary L.
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 79
CORRESPONDENCE ADDRESS:
ADDRESSEE: SHERIDAN ROSS
STREET: 1700 LINCOLN ST., SUITE 3500
CITY: DENVER
STATE: CO
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,443B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary L.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C1-2
TELEPHONE: 303 863-9700
TELEFAX: 303 863-0223
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 64 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-485-443B-75
Query Match 80.9%; Score 38; DB 4; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
Db 35 ILNERWVL 42
RESULT 24
US-08-639-075A-75
Sequence 75, Application US/08639075A
Patent No. 6150125
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,431
FILING DATE: 23-Jan-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/639,075A
FILING DATE: 24-APR-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 64 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-639-075A-75
Query Match 80.9%; Score 38; DB 4; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
Db 35 ILNERWVL 42
RESULT 25
US-09-012-431-75
Sequence 75, Application US/09012431
Patent No. 6180383
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,431
FILING DATE: 23-Jan-1998
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 64 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-09-012-431-75

Query Match 80.9%; Score 38; DB 4; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

Db 35 ILNERWVL 42

RESULT 26
US-09-012-692-75
Sequence 75, Application US/09012692
Patent No. 6214579
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,692
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 64 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-012-692-75

Query Match 80.9%; Score 38; DB 4; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

:::|||||

Db 35 ILNERWVL 42
RESULT 27
US-08-906-613-75
Sequence 75, Application US/08906613
Patent No. 6232096
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 75:
SEQUENCE CHARACTERISTICS:
LENGTH: 64 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-906-613-75

Query Match 80.9%; Score 38; DB 4; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9

Db 35 ILNERWVL 42

RESULT 28
PCT-US95-14442A-75
Sequence 75, Application PC/TUS9514442A
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Heath, Andrew W.
APPLICANT: Yamaka, Miles Yamanaka
APPLICANT: Arfsten, Ann
APPLICANT: Dale, Beverly


```
; APPLICANT: Stiegler, Gary
; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
; TITLE OF INVENTION: INESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14442A
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gary J. Connell
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 64 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-14442A-75

Query Match 80.9%; Score 38; DB 5; Length 64;
Best Local Similarity 75.0%; Pred. No. 2.5;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 29
US-08-485-455D-17
; Sequence 17, Application US/08485455D
; Patent No. 5712143
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary L.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,130C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C1-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-482-130C-17

Query Match 80.9%; Score 38; DB 2; Length 223;
Best Local Similarity 75.0%; Pred. No. 9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C1-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-485-455D-17

Query Match 80.9%; Score 38; DB 1; Length 223;
Best Local Similarity 75.0%; Pred. No. 9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 30
US-08-482-130C-17
; Sequence 17, Application US/08482130C
; Patent No. 5962257
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary L.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,130C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C1-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-482-130C-17

Query Match 80.9%; Score 38; DB 2; Length 223;
Best Local Similarity 75.0%; Pred. No. 9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

QY 2 LVNERWVL 9
:|||||
Db 35 ILNERWVL 42

RESULT 31

US-08-484-211C-17
; Sequence 17, Application US/08484211C
; Patent No. 5972645
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary L.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/484,211C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-484-211C-17

Query Match 80.9%; Score 38; DB 2; Length 223;
Best Local Similarity 75.0%; Pred. No. 9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
:|||||
Db 35 ILNERWVL 42

RESULT 32

US-08-817-795-17
; Sequence 17, Application US/08817795
; Patent No. 6139840
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Frank, Glenn R.
; APPLICANT: Heath, Andrew W.
; APPLICANT: Yamaka, Miles Yamanaka
; APPLICANT: Arfsten, Ann
; APPLICANT: Dale, Beverly
; APPLICANT: Stiegler, Gary

; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
; NUMBER OF SEQUENCES: 119
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/817,795
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14442
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Gary J. Connell
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 223 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-817-795-17

Query Match 80.9%; Score 38; DB 4; Length 223;
Best Local Similarity 75.0%; Pred. No. 9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
:|||||
Db 35 ILNERWVL 42

RESULT 33

US-08-485-443B-17
; Sequence 17, Application US/08485443B
; Patent No. 6146870
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Hunter, Shirley Wu
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary L.
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; TITLE OF INVENTION: MOLECULES AND USES THEREOF
; NUMBER OF SEQUENCES: 79
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SHERIDAN ROSS
; STREET: 1700 LINCOLN ST., SUITE 3500
; CITY: DENVER
; STATE: CO
; COUNTRY: USA
; ZIP: 80203
COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30

;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/485,443B
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Connell, Gary J.
;; REGISTRATION NUMBER: 32,020
;; REFERENCE/DOCKET NUMBER: 2618-25-C1-2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 303 863-9700
;; TELEFAX: 303 863-0223
;; INFORMATION FOR SEQ ID NO: 17:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 223 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-485-443B-17

Query Match 80.9%; Score 38; DB 4; Length 223;
Best Local Similarity 75.0%; Pred. No. 9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 34
PCT-US95-14442A-17
;; Sequence 17, Application PC/TUS9514442A
;; GENERAL INFORMATION:
;; APPLICANT: Grieve, Robert B.
;; APPLICANT: Rushlow, Keith E.
;; APPLICANT: Hunter, Shirley Wu
;; APPLICANT: Frank, Glenn R.
;; APPLICANT: Heath, Andrew W.
;; APPLICANT: Yamaka, Miles Yamanaka
;; APPLICANT: Arfsten, Ann
;; APPLICANT: Dale, Beverly
;; APPLICANT: Stiegler, Gary
;; TITLE OF INVENTION: USE OF PROTEASE INHIBITORS AND
;; TITLE OF INVENTION: PROTEASE VACCINES TO PROTECT ANIMALS FROM FLEA
;; TITLE OF INVENTION: INFESTATION, AND FLEA PROTEASE PROTEINS, NUCLEIC ACID
;; TITLE OF INVENTION: MOLECULES, AND USES THEREOF
;; NUMBER OF SEQUENCES: 119
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sheridan Ross & McIntosh
;; STREET: 1700 Lincoln Street, Suite 3500
;; CITY: Denver
;; STATE: Colorado
;; COUNTRY: USA
;; ZIP: 80203
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US95/14442A
;; FILING DATE:
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Gary J. Connell
;; REGISTRATION NUMBER: 32,020
;; REFERENCE/DOCKET NUMBER:
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (303) 863-9700
;; TELEFAX: (303) 863-0223
;; INFORMATION FOR SEQ ID NO: 17:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 223 amino acids
;; TYPE: amino acid

;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
PCT-US95-14442A-17
Query Match 80.9%; Score 38; DB 5; Length 223;
Best Local Similarity 75.0%; Pred. No. 9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 LVNERWVL 9
Db 35 ILNERWVL 42
RESULT 35
US-08-906-769-17
;; Sequence 17, Application US/08906769
;; Patent No. 6077687
;; GENERAL INFORMATION:
;; APPLICANT: Grieve, Robert B.
;; APPLICANT: Rushlow, Keith E.
;; APPLICANT: Wu Hunter, Shirley
;; APPLICANT: Frank, Glenn R.
;; APPLICANT: Stiegler, Gary
;; APPLICANT: Gaines, Patrick J.
;; APPLICANT: Silver, Gary
;; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
;; TITLE OF INVENTION: MOLECULES AND USES THEREOF
;; NUMBER OF SEQUENCES: 190
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Sheridan Ross & McIntosh
;; STREET: 1700 Lincoln Street, Suite 3500
;; CITY: Denver
;; STATE: Colorado
;; COUNTRY: USA
;; ZIP: 80203
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/906,769
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/639,075
;; FILING DATE: 24-APR-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Connell, Gary J.
;; REGISTRATION NUMBER: 32,020
;; REFERENCE/DOCKET NUMBER: 2618-25-C2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (303) 863-9700
;; TELEFAX: (303) 863-0223
;; INFORMATION FOR SEQ ID NO: 17:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 224 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-08-906-769-17

Query Match 80.9%; Score 38; DB 3; Length 224;
Best Local Similarity 75.0%; Pred. No. 9.1;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 36
US-08-906-616-17

```

; Sequence 17, Application US/08906616
; Patent No. 6121035
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA AMINOPEPTIDASE PROTEINS AND USES THEREOF
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/906,616
; FILING DATE: 05-AUG-1997
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2-3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-906-616-17

Query Match 80.9%; Score 38; DB 3; Length 224;
Best Local Similarity 75.0%; Pred No. 9.1;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 37
US-08-639-075A-17
; Sequence 17, Application US/08639075A
; Patent No. 6150125
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA

```

```

; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/639,075A
; FILING DATE: 24-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-639-075A-17

Query Match 80.9%; Score 38; DB 4; Length 224;
Best Local Similarity 75.0%; Pred No. 9.1;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 38
US-09-012-431-17
; Sequence 17, Application US/09012431
; Patent No. 6180383
; GENERAL INFORMATION:
; APPLICANT: Grieve, Robert B.
; APPLICANT: Rushlow, Keith E.
; APPLICANT: Wu Hunter, Shirley
; APPLICANT: Frank, Glenn R.
; APPLICANT: Stiegler, Gary
; APPLICANT: Gaines, Patrick J.
; APPLICANT: Silver, Gary
; TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
; NUMBER OF SEQUENCES: 190
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, Suite 3500
; CITY: Denver
; STATE: Colorado
; COUNTRY: USA
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/012,431
; FILING DATE: 23-Jan-1998
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/639,075
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2618-25-C2
; TELECOMMUNICATION INFORMATION:

```

TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 17:
US-09-012-431-17

Query Match 80.9%; Score 38; DB 4; Length 224;
Best Local Similarity 75.0%; Pred. No. 9.1;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 39
US-09-012-692-17
Sequence 17, Application US/09012692
Patent No. 6214579
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/012,692
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-012-692-17

Query Match 80.9%; Score 38; DB 4; Length 224;
Best Local Similarity 75.0%; Pred. No. 9.1;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42
RESULT 40
US-08-906-613-17
Sequence 17, Application US/08906613
Patent No. 6232096
GENERAL INFORMATION:
APPLICANT: Grieve, Robert B.
APPLICANT: Rushlow, Keith E.
APPLICANT: Wu Hunter, Shirley
APPLICANT: Frank, Glenn R.
APPLICANT: Stiegler, Gary
APPLICANT: Gaines, Patrick J.
APPLICANT: Silver, Gary
TITLE OF INVENTION: FLEA PROTEASE PROTEINS, NUCLEIC ACID
TITLE OF INVENTION: MOLECULES AND USES THEREOF
NUMBER OF SEQUENCES: 190
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: USA
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/906,613
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/639,075
FILING DATE: 24-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-25-C2
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 224 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-906-613-17

Query Match 80.9%; Score 38; DB 4; Length 224;
Best Local Similarity 75.0%; Pred. No. 9.1;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LVNERWVL 9
Db 35 ILNERWVL 42

RESULT 41
US-08-278-091-8
Sequence 8, Application US/08278091
Patent No. 5506139
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: OOMEN, Raymond P.
APPLICANT: KLEIN, Michel H.

; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/278,091
; FILING DATE: 21-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-371
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-278-091-8

Query Match 80.9%; Score 38; DB 1; Length 232;
; Best Local Similarity 77.8%; Pred. No. 9.4;
; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNPRWVL 9
||||:||||
DB 29 VLVNPRWVL 37

RESULT 42
US-08-483-859-8
; Sequence 8, Application US/08483859
; Patent No. 5656436
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/483,859
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-495 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-483-859-8

Query Match 80.9%; Score 38; DB 1; Length 232;
; Best Local Similarity 77.8%; Pred. No. 9.4;
; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNPRWVL 9
||||:||||
DB 29 VLVNPRWVL 37

RESULT 43
US-08-472-173-8
; Sequence 8, Application US/08472173
; Patent No. 5665353
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,173
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-493 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:

;
; LENGTH: 232 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-472-173-8

Query Match 80.9%; Score 38; DB 1; Length 232;
Best Local Similarity 77.8%; Pred. No. 9.4;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
Db 29 VLVNPKWVL 37

RESULT 44
US-08-487-167-8
; Sequence 8, Application US/08487167
; Patent No. 5869302
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
; TITLE OF INVENTION: Reduced Protease Activity
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/487,167
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-508 MIS:vg
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-487-167-8

Query Match 80.9%; Score 38; DB 2; Length 232;
Best Local Similarity 77.8%; Pred. No. 9.4;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
Db 29 VLVNPKWVL 37

RESULT 45
US-08-482-816-8
; Sequence 8, Application US/08482816
; Patent No. 5935573
; GENERAL INFORMATION:
; APPLICANT: LOOSMORE, Sheena M.
; APPLICANT: YANG, Yan-Ping
; APPLICANT: CHONG, Pele
; APPLICANT: OOMEN, Raymond P.
; APPLICANT: KLEIN, Michel H.
; TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with Reduced Protease
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/482,816
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/296,149
; FILING DATE: 26-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/278,091
; FILING DATE: 21-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I.
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-494 MIS:vg
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-482-816-8

Query Match 80.9%; Score 38; DB 2; Length 232;
Best Local Similarity 77.8%; Pred. No. 9.4;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 VLVNERWVL 9
Db 29 VLVNPKWVL 37

Search completed: November 6, 2002, 12:06:25
Job time: 9.11111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 18.4444 Seconds
(without alignments)
84.413 Million cell updates/sec

Title: US-09-905-083-35
Perfect score: 40
Sequence: 1 LLPLQLILL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_rvirus:*
- 16: sp_bacteriapi:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	95.0	189	11 Q9D8P8	Q9D8P8 mus musculus
2	38	95.0	198	11 Q9CRE4	Q9CRE4 mus musculus
3	38	95.0	201	11 Q9D1M7	Q9D1M7 mus musculus
4	36	90.0	845	16 Q97H76	Q97H76 clostridium
5	35	87.5	208	8 Q21527	Q21527 clethrionom
6	35	87.5	493	16 Q9ZKQ5	Q9ZKQ5 helicobacte
7	35	87.5	496	16 Q25600	Q25600 helicobacte
8	34	85.0	256	11 Q9Z2G5	Q9Z2G5 mus musculus
9	34	85.0	266	11 Q9ESS3	Q9ESS3 mus musculus
10	34	85.0	267	11 Q9R1S4	Q9R1S4 rattus norv
11	34	85.0	267	11 Q35426	Q35426 mus musculus
12	34	85.0	335	6 Q97748	Q97748 bos taurus
13	34	85.0	374	4 Q9NS36	Q9NS36 homo sapien
14	34	85.0	429	4 Q9UFQ4	Q9UFQ4 homo sapien
15	34	85.0	539	16 Q9HTX4	Q9HTX4 pseudomonas
16	34	85.0	761	4 Q9Y2L4	Q9Y2L4 homo sapien

ALIGNMENTS

RESULT 1

ID	Q9D8P8	PRELIMINARY;	PRT;	189 AA.
AC	Q9D8P8;			
DT	01-JUN-2001 (TrEMBLrel. 17, Created)			
DT	01-JUN-2001 (TrEMBLrel. 17, Last sequence update)			
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)			
DE	1110002023RIK PROTEIN.			
GN	1110002023RIK.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN=C57BL/6J; TISSUE=PANCREAS;			
RX	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Alizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,			
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,			
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,			
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mommaerts P.,			
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,			
RA	Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,			
RA	Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,			
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,			
RA	Hayashizaki Y.;			
RA	"Functional annotation of a full-length mouse cDNA collection.";			
RL	Nature 409:685-690(2001).			
DR	EMBL; AK007813; BAB25278.1; -.			
DR	HSSP; P20071; 1PKJ.			

17	34	85.0	867	4	Q9NU50	Q9nu50 homo sapien
18	34	85.0	1080	4	Q94890	Q94890 homo sapien
19	34	85.0	1092	4	Q9V4R1	Q9v4r1 homo sapien
20	34	85.0	1236	2	Q9JPA4	Q9jpa4 rhodocyclus
21	33	82.5	146	17	Q9V2D5	Q9v2d5 pyrococcus
22	33	82.5	149	5	Q9VBD1	Q9vbd1 drosophila
23	33	82.5	173	8	Q9XKN1	Q9xkn1 exaerete fr
24	33	82.5	201	4	Q9NYL4	Q9nyl4 homo sapien
25	33	82.5	259	4	Q9NPP7	Q9npp7 homo sapien
26	33	82.5	277	17	Q9HP13	Q9hp13 halobacteri
27	33	82.5	279	2	Q93C52	Q93c52 serratia ma
28	33	82.5	282	16	Q53979	Q53979 mycobacteri
29	33	82.5	420	2	Q9RL01	Q9rl01 streptomyce
30	32	80.0	100	5	Q9W4Q5	Q9w4q5 drosophila
31	32	80.0	106	12	Q9DW78	Q9dw78 rat cytomeg
32	32	80.0	180	5	Q9NB24	Q9nb24 heliothis z
33	32	80.0	180	5	Q9BRA6	Q9bra6 heliothis z
34	32	80.0	187	10	Q24352	Q24352 silene lati
35	32	80.0	199	17	Q9GYM1	Q9gym1 sulfolobus
36	32	80.0	245	16	Q9RVE6	Q9rve6 deinococcus
37	32	80.0	253	16	Q92DJ1	Q92dj1 listeria in
38	32	80.0	378	6	Q9NIX4	Q9nix4 sus scrofa
39	32	80.0	693	3	Q9CA42	Q9ca42 tricholoma
40	32	80.0	789	2	Q93M73	Q93m73 xanthomonas
41	32	80.0	809	5	Q960C5	Q960c5 drosophila
42	32	80.0	1135	5	Q9VJE6	Q9vje6 drosophila
43	32	80.0	1485	13	P70034	P70034 xenopus lae
44	32	80.0	1873	10	Q9FG11	Q9fg11 arabidopsis
45	31	77.5	96	12	Q9QU08	Q9qu08 tt virus. o

DR MGD: MGI:1913370; 1110002023Rik.
 DR InterPro: IPR001179; FKBP_PP1ase.
 DR Pfam: PF00254; FKBP; 1.
 DR PROSITE: PS00453; FKBP_PP1ase_1; UNKNOWN_1.
 DR PROSITE: PS50059; FKBP_PP1ase_3; 1.
 SQ SEQUENCE 189 AA; 20626 MW; AD9795BF7F1E0582B CRC64;

Query Match 95.0%; Score 38; DB 11; Length 189;
 Best Local Similarity 88.9%; Pred. No. 4.3;
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
 |1111111111|
 Db 7 LLPLQLLLLL 15

RESULT 2
 Q9C9E4 PRELIMINARY; PRT; 198 AA.
 AC Q9C9E4;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 110002023Rik PROTEIN (FRAGMENT).
 GN 1110002023Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Brownstein M.J., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK019132; BAB31559.1; -.
 DR HSSP: P20071; 1FKJ.
 DR MGD: MGI:1913370; 1110002023Rik.
 DR InterPro: IPR001179; FKBP_PP1ase.
 DR PROSITE: PS00453; FKBP_PP1ase_1; UNKNOWN_1.
 DR PROSITE: PS50059; FKBP_PP1ase_3; 1.
 FT NON_TER 198 198
 SQ SEQUENCE 198 AA; 21882 MW; A588345383032972 CRC64;

Query Match 95.0%; Score 38; DB 11; Length 198;
 Best Local Similarity 88.9%; Pred. No. 4.5;
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
 |1111111111|
 Db 7 LLPLQLLLLL 15

RESULT 3
 Q9D1M7 PRELIMINARY; PRT; 201 AA.
 AC Q9D1M7;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DE 110002023Rik PROTEIN.
 GN 1110002023Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Brownstein M.J., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR EMBL: AK019132; BAB31559.1; -.
 DR HSSP: P20071; 1FKJ.
 DR MGD: MGI:1913370; 1110002023Rik.
 DR InterPro: IPR001179; FKBP_PP1ase.
 DR PROSITE: PS00453; FKBP_PP1ase_1; UNKNOWN_1.
 DR PROSITE: PS50059; FKBP_PP1ase_3; 1.
 FT NON_TER 198 198
 SQ SEQUENCE 201 AA; 22137 MW; 94D955C57264BD82 CRC64;

Query Match 95.0%; Score 38; DB 11; Length 201;
 Best Local Similarity 88.9%; Pred. No. 4.6;
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
 |1111111111|
 Db 7 LLPLQLLLLL 15

RESULT 4
 Q97H76 PRELIMINARY; PRT; 845 AA.
 AC Q97H76;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DE 110002023Rik PROTEIN.
 GN 1110002023Rik.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Noelling J., Breton G., Omelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
 RA Bennett G.N., Koonin E.V., Smith D.R.;

RT *Genome sequence and comparative analysis of the solvent-producing
bacterium Clostridium acetobutylicum.*;
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL: AE007714; AAK80095.1; -;
DR InterPro: IPR004014; Cation_ATPase.
DR InterPro: IPR001757; EI-E2_ATPase.
DR InterPro: IPR000695; HATPase.
DR InterPro: IPR001454; Hydrolase.
DR InterPro: IPR000661; Na_H_K_ATPase.
DR Pfam: PF00689; Cation_ATPase_C; 1.
DR Pfam: PF00690; Cation_ATPase_N; 1.
DR Pfam: PF00122; EI-E2_ATPase; 1.
DR Pfam: PF00702; Hydrolase; 1.
DR PRINTS: PR00119; CATATPASE.
DR PRINTS: PR00120; HATPASE.
DR PROSITE: PS00154; ATPASE_E1_E2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 845 AA; 93779 MW; 401293AALFF9D757 CRC64;

Query Match 90.0%; Score 36; DB 16; Length 845;
Best Local Similarity 77.8%; Pred. No. 45;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | | | | | | |
Db 670 LLPLQILL 678

RESULT 5
O21527 PRELIMINARY; PRT; 208 AA.
AC O21527;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3) (FRAGMENT).
GN ND4.
OS Clethrionomys gapperi (Southern red-backed vole).
OG Mithochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;
OC Clethrionomys.
OX NCBI_TaxID=56223;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98152303; PubMed=9491603;
RA Engel S.R., Hogan K.M., Taylor J.F., Davis S.K.;
RT "Molecular systematics and paleobiogeography of the South American
sigmodontine rodents.";
RL Mol. Biol. Evol. 15:35-49(1998).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL: U83808; AAB87168.1; -;
DR InterPro: IPR001750; Oxidored_q1.
DR InterPro: IPR000260; Oxidored_q5_N.
DR Pfam: PF00361; oxidored_q1; 1.
DR Pfam: PF01059; oxidored_q5_N; 1.
DR Mithochondrion; NAD; Oxidoreductase; Ubiquinone.
KW NON_TER 208 208
FT SEQUENCE 208 AA; 23967 MW; 8AF1788697AED6A2 CRC64;

Query Match 87.5%; Score 35; DB 8; Length 208;
Best Local Similarity 77.8%; Pred. No. 19;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | | | | | | |
Db 99 LVPLQILL 107

RESULT 6
Q92KQ5 PRELIMINARY; PRT; 493 AA.
ID Q92KQ5
AC Q92KQ5;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE PUTATIVE.
GN JHP0880.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL: AE001518; AAD06464.1; -;
KW Complete proteome.
SQ SEQUENCE 493 AA; 53597 MW; 7F19A079A844A962 CRC64;

Query Match 87.5%; Score 35; DB 16; Length 493;
Best Local Similarity 88.9%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | | | | | | |
Db 254 LLPLSILL 262

RESULT 7
O25600 PRELIMINARY; PRT; 496 AA.
ID O25600;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONSERVED HYPOTHETICAL INTEGRAL MEMBRANE PROTEIN.
GN HP0946.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=26695 / ATCC 700392;
RC MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kerlavage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., FitzGerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weldman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
pylori.";
RL Nature 388:539-547(1997).
DR EMBL: AE000604; AAD07993.1; -;
DR TIGR: HP0946; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 496 AA; 53928 MW; 3BAA5DD8F14F094B CRC64;

Query Match 87.5%; Score 35; DB 16; Length 496;
Best Local Similarity 88.9%; Pred. No. 44;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | | | | | | |

```
Db 254 LLPLSILL 262

RESULT 8
Q922G5 PRELIMINARY; PRT; 256 AA.
ID Q922G5
AC Q922G5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO X-BOX BINDING PROTEIN 1 (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC008153; AAH08153.1; -.
FT NON_TER 1
SQ SEQUENCE 256 AA; 28651 MW; 5544DB566D7B2620 CRC64;

Query Match 85.0%; Score 34; DB 11; Length 256;
Best Local Similarity 88.9%; Pred. No. 38;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | | | |
Db 177 LLPLQILL 185

RESULT 9
Q9ESS3 PRELIMINARY; PRT; 266 AA.
ID Q9ESS3
AC Q9ESS3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TAX-RESPONSIVE ELEMENT-BINDING PROTEIN 5.
RN TREB5.
GN TREB5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-20363095; PubMed=10907849;
RA Masaki T., Noguchi H., Kobayashi M., Yoshida M., Takamatsu K.;
RT "Isolation and characterization of the gene encoding mouse tax-
responsive element-binding protein (TREB)5.";
RL DNA Res. 7:187-193(2000).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL; AB036745; BAB13793.1; -.
DR InterPro; IPR001871; bzip.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 266 AA; 29582 MW; 4161FFA93B1800A5 CRC64;

Query Match 85.0%; Score 34; DB 11; Length 266;
Best Local Similarity 88.9%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | | | |
Db 187 LLPLQILL 195

RESULT 10
Q9R1S4 PRELIMINARY; PRT; 267 AA.
ID Q9R1S4
AC Q9R1S4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HEPATOCARCINOGENESIS-RELATED TRANSCRIPTION FACTOR (HTF).
GN HTF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=20137507; PubMed=10675042;
RA Kokura K., Kishimoto T., Tamura T.;
RT "Identity between rat htf and human xbp-1 genes: determination of gene
structure, target sequence, and transcription promotion function for
HTF.";
RL Gene 241:297-307(2000).
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL; AB030238; BAA82600.1; -.
DR InterPro; IPR001871; bzip.
DR Pfam; PF00170; bzip; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 267 AA; 29665 MW; B5A58FD3FAA10B4 CRC64;

Query Match 85.0%; Score 34; DB 11; Length 267;
Best Local Similarity 88.9%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | | | |
Db 188 LLPLQILL 196

RESULT 11
Q35426 PRELIMINARY; PRT; 267 AA.
ID Q35426
AC Q35426;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE X-BOX BINDING PROTEIN-1.
GN XBP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Lee C.M., Reddy E.P.;
RT "Sequence Analysis of Murine XBP-1.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL; AF027963; AAB81862.2; -.
DR MGD; MGI:98970; Xbp1.
DR InterPro; IPR001871; bzip.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; BZIP_BASIC; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 267 AA; 29619 MW; 15009E684F6D426F CRC64;

Query Match 85.0%; Score 34; DB 11; Length 267;
Best Local Similarity 88.9%; Pred. No. 39;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | | | |
Db 188 LLPLQILL 196
```

```
RESULT 12
O97748 PRELIMINARY; PRT; 335 AA.
AC O97748;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONGLUTININ
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC MEDLINE=93213261; PubMed=8460993;
RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC MEDLINE=93277452; PubMed=7684896;
RA Lu J., Laursen S., Thiel S., Jensenius J., Reid K.;
RT "The cDNA cloning of conglutinin and identification of liver as a
RT primary site of synthesis of conglutinin in members of the Bovidae.";
RL Biochem. J. 292:157-162(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC MEDLINE=94128104; PubMed=8297370;
RA Kawasaki N., Itoh N., Kawasaki T.;
RT "Gene organization and 5'-flanking region sequence of conglutinin: a
RT C-type mammalian lectin containing a collagen-like domain.";
RL Biochem. Biophys. Res. Commun. 198:597-604(1994).
DR EMBL; D25302; BAA04983.1; JOINED.
DR EMBL; D25296; BAA04983.1; JOINED.
DR EMBL; D25297; BAA04983.1; JOINED.
DR EMBL; D25299; BAA04983.1; JOINED.
DR EMBL; D25300; BAA04983.1; JOINED.
DR EMBL; D25301; BAA04983.1; JOINED.
DR HSSP; P35247; 1B08
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; Lectin_c.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00059; Lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
FT CONFLICT 236 236 E -> V (IN REF. 2).
SQ SEQUENCE 335 AA; 34702 MW; 39D3A30BC76C134C CRC64;

Query Match 85.0%; Score 34; DB 6; Length 335;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
DB 3 LLPLSVLL 11
|||||

RESULT 13
Q9NS36 PRELIMINARY; PRT; 374 AA.
AC Q9NS36;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LYMPHOCYTE MEMBRANE ASSOCIATED PROTEIN.
GN 887.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ma F.-R., Zhu L.-P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043290; AAC02992.1; -.
DR InterPro; IPR002017; Spectrin.
DR SMART; SM00150; SPEC; 1.
SQ SEQUENCE 374 AA; 42187 MW; 01383C0CE42E56D2 CRC64;

Query Match 85.0%; Score 34; DB 4; Length 374;
Best Local Similarity 87.5%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLLLL 9
DB 315 LPLQLLLL 322
|||||

RESULT 14
Q9UFQ4 PRELIMINARY; PRT; 429 AA.
AC Q9UFQ4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 48.2 KDA PROTEIN.
GN DKFZP434H2235.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Poustka A., Klein M., Mewes H.W., Cassenhuber J., Wiemann S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117404; CAB55905.1; -.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00435; spectrin; 2.
DR SMART; SM00150; SPEC; 2.
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 48159 MW; 8BF416B8EEA07C63 CRC64;

Query Match 85.0%; Score 34; DB 4; Length 429;
Best Local Similarity 87.5%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLLLL 9
DB 381 LPLQLLLL 388
|||||

RESULT 15
Q9HTX4 PRELIMINARY; PRT; 539 AA.
AC Q9HTX4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE PROBABLE PERMEASE OF ABC IRON TRANSPORTER.
GN PA5216.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 15692 / PA01;
RL MEDLINE=20437337; PubMed=10984043;
RX
```

```
RESULT 12
O97748 PRELIMINARY; PRT; 335 AA.
AC O97748;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CONGLUTININ
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC MEDLINE=93213261; PubMed=8460993;
RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC MEDLINE=93277452; PubMed=7684896;
RA Lu J., Laursen S., Thiel S., Jensenius J., Reid K.;
RT "The cDNA cloning of conglutinin and identification of liver as a
RT primary site of synthesis of conglutinin in members of the Bovidae.";
RL Biochem. J. 292:157-162(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=LIVER;
RC MEDLINE=94128104; PubMed=8297370;
RA Kawasaki N., Itoh N., Kawasaki T.;
RT "Gene organization and 5'-flanking region sequence of conglutinin: a
RT C-type mammalian lectin containing a collagen-like domain.";
RL Biochem. Biophys. Res. Commun. 198:597-604(1994).
DR EMBL; D25302; BAA04983.1; JOINED.
DR EMBL; D25296; BAA04983.1; JOINED.
DR EMBL; D25297; BAA04983.1; JOINED.
DR EMBL; D25299; BAA04983.1; JOINED.
DR EMBL; D25300; BAA04983.1; JOINED.
DR EMBL; D25301; BAA04983.1; JOINED.
DR HSSP; P35247; 1B08
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; Lectin_c.
DR Pfam; PF01391; Collagen; 2.
DR Pfam; PF00059; Lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE; PS00041; C_TYPE_LLECTIN_2; 1.
FT CONFLICT 236 236 E -> V (IN REF. 2).
SQ SEQUENCE 335 AA; 34702 MW; 39D3A30BC76C134C CRC64;

Query Match 85.0%; Score 34; DB 6; Length 335;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
DB 3 LLPLSVLL 11
|||||

RESULT 13
Q9NS36 PRELIMINARY; PRT; 374 AA.
AC Q9NS36;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LYMPHOCYTE MEMBRANE ASSOCIATED PROTEIN.
GN 887.

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Ma F.-R., Zhu L.-P.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF043290; AAC02992.1; -.
DR InterPro; IPR002017; Spectrin.
DR SMART; SM00150; SPEC; 1.
SQ SEQUENCE 374 AA; 42187 MW; 01383C0CE42E56D2 CRC64;

Query Match 85.0%; Score 34; DB 4; Length 374;
Best Local Similarity 87.5%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLLLL 9
DB 315 LPLQLLLL 322
|||||

RESULT 14
Q9UFQ4 PRELIMINARY; PRT; 429 AA.
AC Q9UFQ4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 48.2 KDA PROTEIN.
GN DKFZP434H2235.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Poustka A., Klein M., Mewes H.W., Cassenhuber J., Wiemann S.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL117404; CAB55905.1; -.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00435; spectrin; 2.
DR SMART; SM00150; SPEC; 2.
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 48159 MW; 8BF416B8EEA07C63 CRC64;

Query Match 85.0%; Score 34; DB 4; Length 429;
Best Local Similarity 87.5%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLLLL 9
DB 381 LPLQLLLL 388
|||||

RESULT 15
Q9HTX4 PRELIMINARY; PRT; 539 AA.
AC Q9HTX4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE PROBABLE PERMEASE OF ABC IRON TRANSPORTER.
GN PA5216.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=ATCC 15692 / PA01;
RL MEDLINE=20437337; PubMed=10984043;
RX
```

```

RA Stover C.K., Pham X.-O.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.,
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL: AE004934; AAG08601.1; -.
DR InterPro: IPR000515; BPD.transp.
DR InterPro: IPR001325; C5_DNA_meth.
DR Pfam: PF00528; BPD.transp. 2.
DR DR PROSITE: PS00402; BPD_TRANSF_INN_MEMBR; UNKNOWN_2.
DR DR PROSITE: PS00095; C5_MTASE_2; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 539 AA; 58555 MW; 6717F52FEEF45108 CRC64;

Query Match 85.0%; Score 34; DB 16; Length 539;
Best Local Similarity 77.8%; Pred. No. 76;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
Db 21 LLPLSVLL 29

RESULT 16
Q9Y2L4 PRELIMINARY; PRT; 761 AA.
AC Q9Y2L4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KIAA1011 PROTEIN (FRAGMENT).
GN KIAA1011.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99246063; PubMed=10231032;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Hirose M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIII.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:63-70(1999).
DR EMBL: AB023228; BAA76855.1; -.
DR InterPro: IPR002017; Spectrin.
DR Pfam: PF00435; spectrin; 4.
DR SMART: SM00150; SPEC; 4.
FT NON_TER 1
SQ SEQUENCE 761 AA; 86614 MW; 162CA2016EB1B26B CRC64;

Query Match 85.0%; Score 34; DB 4; Length 761;
Best Local Similarity 87.5%; Pred. No. 1e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQIILL 9
Db 713 LPLQLLL 720

RESULT 17
Q9NU50 PRELIMINARY; PRT; 867 AA.
AC Q9NU50;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
GN

```

```

DE DJI30E4.2 (KIAA0796) (FRAGMENT).
GN DJI30E4.2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Parker A.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL078582; CAB87586.1; -.
DR InterPro: IPR002017; Spectrin.
DR SMART: SM00150; SPEC; 3.
FT NON_TER 1
SQ SEQUENCE 867 AA; 99247 MW; 1E8DB4F5274ACAF2 CRC64;

Query Match 85.0%; Score 34; DB 4; Length 867;
Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQIILL 9
Db 819 LPLQLLL 826

RESULT 18
O94890 PRELIMINARY; PRT; 1080 AA.
ID O94890
AC O94890;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE KIAA0796 PROTEIN (FRAGMENT).
GN KIAA0796.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K., Suyama M., Kikuno R., Miyajima N., Tanaka A.,
RA Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 5:277-286(1998).
DR EMBL: AB018339; BAA34516.1; -.
DR InterPro: IPR002017; Spectrin.
DR Pfam: PF00435; spectrin; 7.
DR SMART: SM00150; SPEC; 5.
FT NON_TER 1
SQ SEQUENCE 1080 AA; 124093 MW; C2660365BE949CC1 CRC64;

Query Match 85.0%; Score 34; DB 4; Length 1080;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQIILL 9
Db 1032 LPLQLLL 1039

RESULT 19
Q9Y4R1 PRELIMINARY; PRT; 1092 AA.
ID Q9Y4R1
AC Q9Y4R1;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 126.2 KDA PROTEIN.
GN DKF2P434G173.

```

OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RX MEDLINE=21154917; PubMed=11230166;
RA Wiemann S., Weil B., Wellenreuther R., Gassenhuber J., Glassl S.,
RA Ansoerge W., Boecher M., Bloeker H., Bauersachs S., Blum H.,
RA Lauber J., Duesterhoeft A., Beyer A., Koehrer K., Strack N.,
RA Mewes H.W., Ottenwaelder B., Obermaier B., Tampe J., Heubner D.,
RA Wambutt R., Korn B., Klein M., Poustka A.;
RT "Towards a Catalog of Human Genes and Proteins: Sequencing and
RT Analysis of 500 Novel Complete Protein Coding Human CDNAS.";
RL Genome Res. 11:422-435(2001).
DR EMBL; AL080133; CAB45729.1; -.
DR InterPro; IPR000488; Death.
DR InterPro; IPR002017; Spectrin.
DR Pfam; PF00435; spectrin; 7.
DR SMART; SM00005; spectrin; 7.
DR SMART; SM00150; SPECT; 7.
KW Hypothetical protein.
SQ SEQUENCE 1092 AA; 126211 MW; 1D7331C888D5D439 CRC64;

Query Match 85.0%; Score 34; DB 4; Length 1092;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
Db 104 LPLQLLL 1051
|||||
RESULT 20
Q9JPA4 PRELIMINARY; PRT; 1236 AA.
AC Q9JPA4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MG PROTOPORPHYRIN METHYL TRANSFERASE.
GN BCHH.
OS Rhodocyclus gelatinosus (Rhodopseudomonas gelatinosa).
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC Rubrivivax.
OX NCBI_TaxID=28068;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL144;
RA Nagashima K.V., Shimada K., Matsuura K.;
RT "Phylogenetic analysis of photosynthetic genes of Rhodocyclus
RT gelatinosus: Possibility of horizontal gene transfer in purple
RT bacteria.";
RL Photosyn. Res. 36:185-191(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=IL144;
RX MEDLINE=94132007; PubMed=8300574;
RA Nagashima K.V., Matsuura K., Ohyama S., Shimada K.;
RT "Primary structure and transcription of genes encoding B870 and
RT photosynthetic reaction center apoproteins from Rubrivivax
RT gelatinosus.";
RL J. Biol. Chem. 269:2477-2484 (1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=IL144;
RA Igarashi N., Shimada K., Matsuura K., Nagashima K.V.;
RT "Photosynthetic gene cluster in purple bacterium, Rubrivivax
RT gelatinosus.";
RL (in) Garab G. (eds.);
RL Photosynthesis:
RL mechanisms and effects (Proceedings of the 11th international congress

RL on photosynthesis), pp.4:2889-2892, Kluwer Academic Publishers,
RL Dordrecht (1999).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=IL144;
RX MEDLINE=20031519; PubMed=10563807;
RA Menin L., Yoshida M., Jaquinod M., Nagashima K.V., Matsuura K.,
RA Parot P., Vermeiglio A.;
RT "Park aerobic growth conditions induce the synthesis of a high
RT midpoint potential cytochrome c8 in the photosynthetic bacterium
RT Rubrivivax gelatinosus.";
RL Biochemistry 38:15238-15244(1999).
DR EMBL; AB034704; BAA94057.1; -.
DR InterPro; IPR003672; CobN/Mg-chelatase.
DR Pfam; PF02514; cobN-Mg_chel; 1.
KW Transferase.
SQ SEQUENCE 1236 AA; 134729 MW; 84051C045638520C CRC64;

Query Match 85.0%; Score 34; DB 2; Length 1236;
Best Local Similarity 88.9%; Pred. No. 1.6e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
Db 946 LLPLQIKLL 954
|||||
RESULT 21
Q9V2D5 PRELIMINARY; PRT; 146 AA.
AC Q9V2D5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 16.1 KDA PROTEIN.
GN PAB0088.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution.";
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248283; CAB49063.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 146 AA; 16092 MW; 7182941371258C1F CRC64;

Query Match 82.5%; Score 33; DB 17; Length 146;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
Db 68 LLPLQIIIV 76
|||||
RESULT 22
Q9V8D1 PRELIMINARY; PRT; 149 AA.
AC Q9V8D1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG14498 PROTEIN.
GN CG14498.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;

```

RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BERKELEY;
RX  MEDLINE=20196006; PubMed=10731132;
RA  Adams M.D., Celnikier S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Anantides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA  Arif J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA  Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA  Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA  Borkov D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA  Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA  Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA  de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA  Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA  Durbin K.J., Evangelista C.C., Ferraz C., Ferrier S., Fleischmann W.,
RA  Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA  Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA  Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA  Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA  Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA  Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA  Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA  Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA  Minkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA  Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA  Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA  Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA  Reinart K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA  Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA  Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA  Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA  Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA  Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA  Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA  Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA  Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT  "The genome sequence of Drosophila melanogaster.";
RL  Science 287:2185-2195(2000).
DR  EMBL: AE003800; AAF57736.1; -.
DR  FlyBase: FBgn0034309; CGI4498.
DR  InterPro: IPR000847; HTH_LYSR.
DR  PROSITE: PS00044; HTH_LYSR_FAMILY; UNKNOWN1.
SQ  SEQUENCE 149 AA; 14815 MW; 7E7A95537FEB00A6 CRC64;

Query Match 82.5%; Score 33; DB 5; Length 149;
Best Local Similarity 66.7%; Pred. No. 36;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQLILL 9
Db 41 MLPLELLLL 49

RESULT 23
Q9XKN1 PRELIMINARY; PRT; 173 AA.
AC Q9XKN1
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYTOCHROME B (FRAGMENT)
GN CYTB.
OS Exaerete frontalis.
OG Mitochondrion.
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata;
OC Apoidea; Apidae; Exaerete.
OX NCBI_TaxID=60905;
RN [1]

```

```

RP  SEQUENCE FROM N.A.
RA  Koulianos S., Schmid-Hempel R., Roubik D., Schmid-Hempel P.;
RT  "Relationships within the Apinae.";
RL  Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
CC  -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC  COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC  RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC  COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC  -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC  BOUND TO THE PROTEIN (BY SIMILARITY).
CC  -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC  CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
DR  EMBL: AF002729; AAD29079.1; -.
DR  InterPro: IPR000179; Cyt_b_b6.
DR  Pfam: PF00033; cytochrome_b_N; 1.
KW  Electron transport; Heme; Mitochondrion; Respiratory chain;
KW  Transmembrane.
FT  NON_TER 1 173
FT  NON_TER 173 173
SQ  SEQUENCE 173 AA; 20616 MW; 7162085525272D26 CRC64;

Query Match 82.5%; Score 33; DB 8; Length 173;
Best Local Similarity 77.8%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQLILL 9
Db 96 LLPLMLILL 104

RESULT 24
Q9NYL4 PRELIMINARY; PRT; 201 AA.
AC Q9NYL4
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE FK506 BINDING PROTEIN PRECURSOR.
GN FKBP19.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP  SEQUENCE FROM N.A.
RA  Rulten S., Kay J.E., Robinson C.;
RT  "Identification of Novel FKBP genes.";
RL  Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AF238079; AAF63478.1; -.
DR  HSSP: Q00688; 1PBK.
DR  InterPro: IPR001179; FKBP_PPase.
DR  Pfam: PF00254; FKBP; 1.
DR  PROSITE: PS00453; FKBP_PPase_1; UNKNOWN1.
DR  PROSITE: PS00509; FKBP_PPase_3; 1.
KW  Signal.
FT  SIGNAL 1 25
FT  SIGNAL 26 201
FT  CHAIN 26 201
FT  CHAIN 201 AA; 22180 MW; 586E430B9D2D0A2 CRC64;
SQ  SEQUENCE 201 AA; 22180 MW; 586E430B9D2D0A2 CRC64;

Query Match 82.5%; Score 33; DB 4; Length 201;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQLILL 9
Db 7 LLPLHLLLL 15

RESULT 25
Q9NPP7 PRELIMINARY; PRT; 259 AA.
ID Q9NPP7
AC Q9NPP7;

```



```
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LDL INDUCED ENDOTHELIAL CELL PROTEIN (FRAGMENT).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Andreu N., Estivill X., Escarceller M., Sunoy L.;
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RA Auffray C., Ansorge W., Ballabio A., Estivill X., Gibson K.,
RA Lehrach H., Poustka A., Lundeberg J.;
RL "The European IMAGE consortium for integrated Molecular analysis of
RT human gene transcripts.";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL359054; CAB94389.1; -
DR HSP: P08047; ISP1.
DR InterPro; IPR002965; P_rich_extensn.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF00096; zf-C2H2; 2.
DR PRINTS; PR01217; PRICHEXTENS.
DR SMART; SM00355; Znf.C2H2; 3.
DR PROSITE; PS00028; ZINC_FINGER_C2H2.1; 3.
DR PROSITE; PS0157; ZINC_FINGER_C2H2.2; 3.
KW DNA-binding; Metal-binding; Zinc-finger.
FT NON_TER
FT 1
SQ SEQUENCE 259 AA; 27171 MW; 309249364112FC70 CRC64;

Query Match 82.5%; Score 33; DB 4; Length 259;
Best Local Similarity 87.5%; Pred. No. 61;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LPLQILL 9
Db 84 LPLQMLL 91

RESULT 26
Q9HPI3 PRELIMINARY; PRT; 277 AA.
AC Q9HPI3;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE CYTOCHROME AA3 CONTROLLING PROTEIN.
GN CCP OR VNG1623G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Fohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005072; AAG19884.1; -
KW Complete proteome.
SQ SEQUENCE 277 AA; 29053 MW; B053123766E274B1 CRC64;

Query Match 82.5%; Score 33; DB 17; Length 277;
Best Local Similarity 75.0%; Pred. No. 65;
```

```
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLOILL 8
Db 98 LLPLOVIL 105

RESULT 27
Q93C52 PRELIMINARY; PRT; 279 AA.
ID Q93C52;
AC Q93C52;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE TOXR (FRAGMENT).
DE TOXR.
GN Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ECB 26;
RA Bhattacharyya T., Bhadra R.K., Nair G.B.;
RT "The toxR gene from ECB 26, a Serratia marcescens strain.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF414370; AAL09685.1; -
FT NON_TER
FT 1
SQ SEQUENCE 279 AA; 30822 MW; 1D5435712AE52492 CRC64;

Query Match 82.5%; Score 33; DB 2; Length 279;
Best Local Similarity 77.8%; Pred. No. 65;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLOILL 9
Db 182 LLPLAVLL 190

RESULT 28
O53979 PRELIMINARY; PRT; 282 AA.
ID O53979;
AC O53979;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 30.6 KDA PROTEIN.
GN RV1978 OR MTV051.16.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=H37RV;
MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
DR EMBL; AL022073; CAA17851.1; -
DR TuberculList; RV1978; -
DR InterPro; IPR000051; SAM_bind.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 282 AA; 30584 MW; 3CA18BD208951D1E CRC64;
```

```
Query Match      82.5%; Score 33; DB 16; Length 282;
Best Local Similarity 77.8%; Pred. No. 66;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 LLPLQIILL 9
Db 197 LLPLHLULL 205

RESULT 29
Q9RL01
ID Q9RL01 PRELIMINARY; PRT; 420 AA.
AC Q9RL01;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE TRANSMEMBRANE TRANSPORT PROTEIN.
GN SC5G9.26C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA James K.D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; Pubmed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kanashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996)
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE SUGAR TRANSPORTER FAMILY.
DR EMBL; ALI17385; CAB55674.1; -
DR InterPro; IPR003662; sub.transporter.
DR Pfam; PF00083; sugar_tr; 1.
KW Transmembrane.
SQ SEQUENCE 420 AA; 42250 MW; 87D5EB267AFD8A8D CRC64;

Query Match      82.5%; Score 33; DB 2; Length 420;
Best Local Similarity 77.8%; Pred. No. 96;
Matches 7; Conservative 1; Mismatches 0; Gaps 0;

Qy 1 LLPLQIILL 9
Db 218 LLPLTVLL 226

RESULT 30
Q9W4Q5
ID Q9W4Q5 PRELIMINARY; PRT; 100 AA.
AC Q9W4Q5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE CG12462 PROTEIN.
GN CG12462.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAASTRICHT;
RX MEDLINE=20366325; Pubmed=10906222;
RA Vink C., Beuken E., Bruggeman C.A.;
RT "Complete DNA sequence of the rat cytomegalovirus genome.";
```

```
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; Pubmed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abill J.F., Agbayani A., An H.-J., Andrews-Pfankoch C., Baldwin D.,
RA Baliew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003428; AAF45891.1; -
DR FlyBase; FBgn0029675; CG12462.
SQ SEQUENCE 100 AA; 11465 MW; 5195CC6061D622E8 CRC64;

Query Match      80.0%; Score 32; DB 5; Length 100;
Best Local Similarity 77.8%; Pred. No. 40;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQIILL 9
Db 13 LLPLMULL 21

RESULT 31
Q9DW78
ID Q9DW78 PRELIMINARY; PRT; 106 AA.
AC Q9DW78;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE PR119.3.
GN R119.3.
OS Rat cytomegalovirus (strain Maastricht).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Muromegalovirus.
OX NCBI_TaxID=97900;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAASTRICHT;
RX MEDLINE=20366325; Pubmed=10906222;
RA Vink C., Beuken E., Bruggeman C.A.;
RT "Complete DNA sequence of the rat cytomegalovirus genome.";
```

```

RL J. Virol. 74:7656-7665(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAASTRICHT;
RX MEDLINE=20473137; PubMed=11018281;
RA Gruijthuisen Y.K., Beuken E., Bruggeman C.A., Vink C.;
RT "Rat cytomegalovirus R89 is a highly conserved gene which expresses a
RT spliced transcript.";
RL Virus Res. 69:119-130(2000).
DR EMBL: AF232689; AAF99212.1; -.
SQ SEQUENCE 106 AA; 11347 MW; 3E616583EBBBA455 CRC64;

Query Match 80.0%; Score 32; DB 12; Length 106;
Best Local Similarity 77.8%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
Db 93 LLPLATILL 101
|||||:|

RESULT 32
Q9NB24 PRELIMINARY; PRT; 180 AA.
AC Q9NB24;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ACYL-COA DELTA-9-LIKE DESATURASE (FRAGMENT).
GN PGDS3.
OS Heliothis zea (Corn earworm) (Bollworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuoidea; Noctuidae; Heliothinae; Helicoverpa.
OX NCBI_TaxID=7113;

RP SEQUENCE FROM N.A.
RC TISSUE=PEROMONE GLAND;
RX MEDLINE=21376154; PubMed=11483431;
RA Rosenfield C., You K.M., Marsella-Herrick P., Roelofs W.L.,
RA Knipple D.C.;
RT "Structural and functional conservation and divergence among acyl-CoA
RT desaturases of two noctuid species, the corn earworm, Helicoverpa
RT zea(l), and the cabbage looper, Trichoplusia ni.";
RL Insect Biochem. Mol. Biol. 31:949-964(2001).
DR EMBL: AF272344; AAF81789.1; -.
DR InterPro: IPR001522; Desaturase.
DR Pfam: PF01069; Desaturase; 1.
DR PRINTS; PR00075; FACDSDSATRASE.
DR ProDom: PD002221; Desaturase; 1.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 21315 MW; E33848743EDF69AC CRC64;

Query Match 80.0%; Score 32; DB 5; Length 180;
Best Local Similarity 87.5%; Pred. No. 69;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQIILL 9
Db 14 LPLRIILL 21
|||||:|

RESULT 33
Q9BKA6 PRELIMINARY; PRT; 180 AA.
AC Q9BKA6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ACYL-COA DELTA-9-LIKE DESATURASE (FRAGMENT).
GN PGDS3.
OS Heliothis zea (Corn earworm) (Bollworm).

```

```

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuoidea; Noctuidae; Heliothinae; Helicoverpa.
OX NCBI_TaxID=7113;

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21376154; PubMed=11483431;
RA Rosenfield C., You K.M., Marsella-Herrick P., Roelofs W.L.,
RA Knipple D.C.;
RT "Structural and functional conservation and divergence among acyl-CoA
RT desaturases of two noctuid species, the corn earworm, Helicoverpa
RT zea(l), and the cabbage looper, Trichoplusia ni.";
RL Insect Biochem. Mol. Biol. 31:949-964(2001).
DR EMBL: AF297109; AAK21861.1; -.
DR InterPro: IPR001522; Desaturase.
DR Pfam: PF01069; Desaturase; 1.
DR PRINTS; PR00075; FACDSDSATRASE.
DR ProDom: PD002221; Desaturase; 1.
FT NON_TER 1
FT NON_TER 180
SQ SEQUENCE 180 AA; 21301 MW; 163D1D74386F6F1A CRC64;

Query Match 80.0%; Score 32; DB 5; Length 180;
Best Local Similarity 87.5%; Pred. No. 69;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQIILL 9
Db 14 LPLRIILL 21
|||||:|

RESULT 34
O24352 PRELIMINARY; PRT; 187 AA.
AC O24352;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MEN-4 PROTEIN.
GN MEN-4.
OS Silene latifolia.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Caryophyllaceae; Silene.
OX NCBI_TaxID=37657;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MALE FLOWER;
RX MEDLINE=97377129; PubMed=9232878;
RA Scutt C.P., Li Y., Robertson S.E., Willis M.E., Gilmartin P.M.;
RT "Sex determination in Silene latifolia: Y chromosome- and Ustilago
RT violacea-mediated effects during dioecious flower development.";
RL Plant Physiol. 114:969-979(1997).
DR EMBL: Y08776; CAA70029.1; -.
SQ SEQUENCE 187 AA; 18357 MW; FAF6GB5E7F7DFEE1 CRC64;

Query Match 80.0%; Score 32; DB 10; Length 187;
Best Local Similarity 66.7%; Pred. No. 72;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
Db 9 LLPLSVLLI 17
|||||:|

RESULT 35
Q96YM1 PRELIMINARY; PRT; 199 AA.
AC Q96YM1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN ST2153.

```

```

GN ST2153.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka K., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT Crenarchaeon, Sulfolobus tokodaii strain7.";
RL DNA Res. 8:123-140(2001).
DR EMBL: AF000989; BAB67256.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 199 AA; 22339 MW; BB27603CC58DFFC4 CRC64;

Query Match 80.0%; Score 32; DB 17; Length 199;
Best Local Similarity 75.0%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
   :|||||:
DB 3 IPLQILL 10

RESULT 36
Q9RVE6
ID Q9RVE6 PRELIMINARY; PRT; 245 AA.
AC Q9RVE6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 25.9 KDA PROTEIN.
GN DR1083.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R1;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Mofat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1571-1577(1999).
DR EMBL: AE001958; AAF10660.1; -.
DR TIGR: DR1083; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 245 AA; 25889 MW; FEDF6FF58C81AEC3 CRC64;

Query Match 80.0%; Score 32; DB 16; Length 245;
Best Local Similarity 77.8%; Pred. No. 93;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
   :|||||:
DB 15 LLPLALL 23

RESULT 37
Q92DJ1
ID Q92DJ1 PRELIMINARY; PRT; 253 AA.

```

```

AC Q92DJ1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LIN0822 PROTEIN.
GN LIN0822.
OS Listeria innocua.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Krefz J., Kuhn M., Kunst F., Kurapkut G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluster T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RT "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL: AL596166; CAC96054.1; -.
DR Listlist: LIN00822; -.
KW Complete proteome.
SQ SEQUENCE 253 AA; 27915 MW; E618DC96B782C05 CRC64;

Query Match 80.0%; Score 32; DB 16; Length 253;
Best Local Similarity 66.7%; Pred. No. 96;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
   :|||||:
DB 232 LLPLLVLL 240

RESULT 38
Q9N1X4
ID Q9N1X4 PRELIMINARY; PRT; 378 AA.
AC Q9N1X4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE LUNG SURFACTANT PROTEIN D PRECURSOR.
GN SFTPD.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20109098; PubMed=10640760;
RA van Eijk M., Haagsman H.P., Skinner T., Archibald A.L., Reid K.B.M.,
RA Lawson P.R.;
RT "Porcine Lung Surfactant Protein D (SP-D): cDNA cloning, chromosomal
RT localisation and tissue distribution.";
RL J. Immunol. 164:1442-1450(2000).
DR EMBL: AF132496; AAF22145.2; -.
DR HSSP: P35247; I808.
DR InterPro: IPR000087; Collagen.
DR InterPro: IPR001304; lectin_c.
DR Pfam: PF01391; Collagen; 3.
DR Pfam: PF00059; lectin_c; 1.
DR SMART: SM00034; CLECT; 1.
DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
DR PROSITE: PSS00041; C_TYPE_LLECTIN_2; 1.
KW Signal.
FT SIGNAL 1 20 POTENTIAL.

```

```

FT CHAIN          21      378      LUNG SURFACTANT PROTEIN D.
SQ SEQUENCE      378 AA; 37986 MW; 3504E8C1E56C341D CRC64;

Query Match      80.0%; Score 32; DB 6; Length 378;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
    |||| :||:
Db 3 LLPLSVLIL 11

RESULT 39
Q9C4A2 PRELIMINARY; PRT; 693 AA.
AC Q9C4A2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE POL-LIKE PROTEIN POL-2.
OS Tricholoma matsutake.
OC Eukaryota; Fungi; Basidiomycota; Hymenomycetes; Homobasidiomycetes;
OC Agaricales; Tricholomataceae; Tricholoma.
OX NCBI_TaxID=40145;
RN [1]
RP SEQUENCE FROM N.A.
RA Murata H., Miyazaki Y., Yamada A.;
RT "marY2N, a LINE-like non-long terminal repeat (non-LTR) retroelement
RL from the ectomycorrhizal homobasidiomycete Tricholoma matsutake.";
RL Biosci. Biotechnol. Biochem. 0:0-0(2001).
DR EMBL; AB047280; BAB32470.1; -.
DR InterPro; IPR000477; RVTse.
DR Pfam; PF00078; rvt; 1.
KW RNA-directed DNA polymerase.
SQ SEQUENCE 693 AA; 76528 MW; 137015E7E7EB7BE9 CRC64;

Query Match      80.0%; Score 32; DB 3; Length 693;
Best Local Similarity 75.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 8
    ||||:||||
Db 493 LLPMQILL 500

RESULT 40
Q93M73 PRELIMINARY; PRT; 789 AA.
AC Q93M73;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE DIPEPTIDYL PEPTIDASE.
OS Xanthomonas oryzae pv. oryzae.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xanthomonas.
OX NCBI_TaxID=64187;
RN [1]
RP SEQUENCE FROM N.A.
RA Goel A.K., Rajagopal L., Nagesh N., Sonti R.V.;
RT "Xanthomonadin biosynthetic gene cluster from Xanthomonas oryzae pv.
RL oryzae.";
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY010120; AAK82926.1; -.
SQ SEQUENCE 789 AA; 88112 MW; 6D38F23EBAD9A008 CRC64;

Query Match      80.0%; Score 32; DB 2; Length 789;
Best Local Similarity 77.8%; Pred. No. 2.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
    ||||:||||
Db 4 LLPLTLLL 12

```

```

RESULT 41
Q960C5 PRELIMINARY; PRT; 809 AA.
AC Q960C5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SD07737P.
GN CG6860.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY052126; AAK93550.1; -.
SQ SEQUENCE 809 AA; 87105 MW; 663DEDF3C70661E9 CRC64;

Query Match      80.0%; Score 32; DB 5; Length 809;
Best Local Similarity 75.0%; Pred. No. 2.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
    ||||:||||
Db 167 LPLQVLLV 174

RESULT 42
Q9VJE6 PRELIMINARY; PRT; 1135 AA.
AC Q9VJE6;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG6860 PROTEIN.
GN CG6860.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananatiades P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.I., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

```

```
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kianos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Rector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.: ".
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
DR EMBL: AE003654; AAF53603.1; -.
DR FlyBase: FBgn0032633; CG6860.
DR InterPro: IPR001715; Calponin_hom.
DR InterPro: IPR003247; CH_type.
DR InterPro: IPR000595; cNMP_binding.
DR InterPro: IPR001611; LRR.
DR InterPro: IPR003592; LRR_out.
DR InterPro: IPR003591; LRR_type.
DR Pfam: PF00560; LRR_7.
DR PRINTS: PR00019; LEURICHRPT.
DR ProDom: PD001527; CH_type; 1.
DR SMART: SM00033; CH; 1.
DR SMART: SM00370; LRR; 2.
DR SMART: SM00369; LRR_TYP; 1.
DR PROSITE: PS50021; CH; 1.
DR PROSITE: PS50042; cNMP_BINDING_3; 1.
SQ SEQUENCE 1135 AA; 124195 MW; 373E884D6253600A CRC64;

Query Match 80.0%; Score 32; DB 5; Length 1135;
Best Local Similarity 75.0%; Pred. No. 3.9e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LPLQLILL 9
Db 167 LPLQLLV 174

RESULT 43
P70034
ID P70034 PRELIMINARY; PRT; 1485 AA.
AC P70034;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYSTIC FIBROSIS TRANSMEMBRANE CONDUCTANCE REGULATOR (FRAGMENT).
GN CFTR.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96411723; PubMed=8810276;
RA Price M.P., Ishihara H., Sheppard D.N., Welsh M.J.:
RT "Function of Xenopus cystic fibrosis transmembrane conductance to
RT regulator (CFTR) Cl channels and use of human-Xenopus chimeras to
RT investigate the pore properties of CFTR."
RL J. Biol. Chem. 271:25184-25191(1996).
DR EMBL: U60209; AAC60023.1; -.
DR HSP: P13569; 1NBD.
DR InterPro: IPR003593; AAA.
DR InterPro: IPR001140; ABC_transporter_tmern.
DR InterPro: IPR003439; ABC_transportr.
```

```
DR InterPro: IPR001687; ATP_GTP_A.
DR Pfam: PF00664; ABC_membrane; 2.
DR Pfam: PF00005; ABC_tran; 2.
DR SMART: SM00382; AAA; 2.
DR PROSITE: PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW ATP-binding; Transmembrane.
FT NON_TER 1485 1485
SQ SEQUENCE 1485 AA; 168502 MW; 0BEAF0D2264D5A24 CRC64;

Query Match 80.0%; Score 32; DB 13; Length 1485;
Best Local Similarity 66.7%; Pred. No. 5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQLILL 9
Db 204 IAPLOVLLL 212

RESULT 44
Q9FGI1
ID Q9FGI1 PRELIMINARY; PRT; 1873 AA.
AC Q9FGI1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GENOMIC DNA, CHROMOSOME 5, pl CLONE:MQJ2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RA Kaneko T., Katoh T., Asamizu E., Sato S., Nakamura Y., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. XI.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB025632; BAB10256.1; -.
DR EMBL: AB019228; BAB10256.1; JOINED.
DR InterPro: IPR001841; Znf_ring.
DR SMART: SM00184; RING; 1.
SQ SEQUENCE 1873 AA; 208743 MW; 777F4E6E620854EA CRC64;

Query Match 80.0%; Score 32; DB 10; Length 1873;
Best Local Similarity 75.0%; Pred. No. 6.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQLILL 8
Db 1194 LLPLEVLL 1201

RESULT 45
Q9QU08
ID Q9QU08 PRELIMINARY; PRT; 96 AA.
AC Q9QU08;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ORF2 PROTEIN (FRAGMENT).
OS TT virus.
OC Viruses; ssDNA viruses; unclassified ssDNA viruses.
OX NCBI_TaxID=68887;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TTV BK1024;
RX MEDLINE=20129505; PubMed=10664406;
RA Mulyanto, Hijikata M., Matsushita M., Ingkokusmo G., Widjaya A.,
RA Sumarsidi D., Kanai K., Ohta Y., Mishiro S.;
RA "TT virus (TTV) genotypes in native and non-native prostitutes of
RT Irian Jaya, Indonesia: implication for non-occupational
RT transmission.";
```

RL Arch. Virol. 145:63-72(2000).
 DR EMBL: AB031685; BAA86148.1; -.
 DR InterPro: IPR004118; TT_ORF2.
 DR Pfam: PF02957; TT_ORF2; 1.
 FT NON_TER 1
 FT NON_TER 96
 SQ SEQUENCE 96 AA; 11069 MW; FC006B668AC1F8F1 CRC64;

Query Match 77.5%; Score 31; DB 12; Length 96;
 Best Local Similarity 87.5%; Pred. No. 62;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 8
 |||||
 Db 13 LLPLQTLL 20

Search completed: November 6, 2002, 12:12:07
 Job time : 21.4444 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 5.33333 Seconds
(without alignments)
65.339 Million cell updates/sec

Title: US-09-905-083-35
Sequence: 1 LLPLQILLL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	40	100.0	253	1	KLK7_HUMAN
2	38	95.0	361	1	IHA_TRIVU
3	34	85.0	369	1	PSPD_BOVIN
4	34	85.0	371	1	CONG_BOVIN
5	33	82.5	294	1	TOXR_VIBCH
6	33	82.5	435	1	YA06_HUMAN
7	33	82.5	500	1	YA04_HUMAN
8	33	82.5	653	1	APPT1_MOUSE
9	32	80.0	247	1	MCT1_PAPHA
10	32	80.0	256	1	LPSB_PERAM
11	32	80.0	394	1	ARAJ_ECOLI
12	32	80.0	412	1	SEPL_HUMAN
13	32	80.0	492	1	MM11_MOUSE
14	32	80.0	606	1	MM17_HUMAN
15	32	80.0	1306	1	ACE_HUMAN
16	32	80.0	1485	1	CFTR_XENLA
17	31	77.5	209	1	EFA2_MOUSE
18	31	77.5	213	1	EFA2_HUMAN
19	31	77.5	327	1	PPA5_RAT
20	31	77.5	328	1	IBP2_HUMAN
21	31	77.5	428	1	YFJD_ECOLI
22	31	77.5	461	1	NCBI_HUMAN
23	31	77.5	480	1	HRAL_HUMAN
24	31	77.5	530	1	PVR2_MOUSE
25	31	77.5	650	1	APPI_HUMAN
26	31	77.5	666	1	NX2B_HUMAN
27	31	77.5	700	1	PTPE_HUMAN
28	31	77.5	1007	1	CHC2_HUMAN
29	31	77.5	3298	1	PC16_HUMAN
30	30	75.0	128	1	POL_FLY
31	30	75.0	242	1	CD8A_BOVIN
32	30	75.0	259	1	YCBC_ECOLI
33	30	75.0	342	1	VG74_KSHV

34	30	75.0	370	1	GPRA_HUMAN	P49683 homo sapien
35	30	75.0	478	1	BM3B_HUMAN	P55107 homo sapien
36	30	75.0	481	1	X389_CLOPE	P30818 clostridium
37	30	75.0	487	1	TTDT_ECOLI	P39414 escherichia
38	30	75.0	492	1	N04M_CHOCH	P48915 chondrus cr
39	30	75.0	537	1	IL2B_RAT	P26896 rattus norv
40	30	75.0	538	1	YA94_HUMAN	O9upq8 homo sapien
41	30	75.0	658	1	SOHC_BRAJA	P54924 bradyrhizob
42	30	75.0	810	1	GARN_DROME	P54362 drosophila
43	30	75.0	2499	1	MPRI_BOVIN	P08169 bos taurus
44	30	75.0	2820	1	NFI_RAT	P97526 rattus norv
45	30	75.0	2839	1	NFI_HUMAN	P21359 homo sapien

ALIGNMENTS

RESULT 1
KLK7_HUMAN
ID KLK7_HUMAN STANDARD; PRT; 253 AA.
AC P49852;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic
enzyme) (hSCCE).
DE KLK7 OR PRSS6 OR SCCE.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.
RC TISSUE=Skin;
RX MEDLINE=94308225; PubMed=8034709;
RA Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A.,
Egelrud T.;
RT "Cloning, expression, and characterization of stratum corneum
chymotryptic enzyme. A skin-specific human serine proteinase.";
RL J. Biol. Chem. 269:19420-19426(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Keratinocytes;
RA Yousef G.M., Scorrilas A., Diamandis E.P.;
RT "Molecular characterization, mapping and tissue expression of the
human stratum corneum chymotryptic enzyme gene.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=20510030; PubMed=11054574;
RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
Moss P., Paepel B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene
cluster located in chromosome 19q13 region.";
RL Gene 257:119-130(2000).
RN [4]
RP CHARACTERIZATION.
RX MEDLINE=95314630; PubMed=7794273;
RA Skjott A., Stroemqvist M., Egelrud T.;
RT "Primary substrate specificity of recombinant human stratum corneum
chymotryptic enzyme.";
RL Biochem. Biophys. Res. Commun. 211:586-589(1995).
CC -1- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE
STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS
SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID
RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE
CLEAVES INSULIN B CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE-
1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION
OF PRECURSORS TO INFLAMMATORY CYTOKINES.
CC -1- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS
EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE
ALSO SEEN IN THE BRAIN AND KIDNEY.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

```

CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L33404; AAC37551.1; -.
CC EMBL: AF166330; AAD49718.1; -.
CC EMBL: AF243527; AAG33360.1; -.
CC HSSP: P00763; IDPO.
CC MEROPS: S01.300; -.
CC MIM: 604438; -.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00020; TRYSP_SPC; 1.
CC PROSITE: PS02440; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SRP; 1.
CC KW Hydrolyase; Serine protease; zymogen; Glycoprotein; Signal.
CC FT SIGNAL 1 22
CC FT PROPEP 23 29 ACTIVATION PEPTIDE..
CC FT CHAIN 30 253 KALLIKREIN 7.
CC FT ACT_SITE 70 70 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
CC FT DISULFID 36 137 BY SIMILARITY.
CC FT DISULFID 55 71 BY SIMILARITY.
CC FT DISULFID 137 239 BY SIMILARITY.
CC FT DISULFID 144 211 BY SIMILARITY.
CC FT DISULFID 176 190 BY SIMILARITY.
CC FT DISULFID 201 226 BY SIMILARITY.
CC FT CARBOHYD 246 246 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SQ SEQUENCE 253 AA; 27525 MW; 2D68B6B15A76A668 CRC64;
Query Match 100.0%; Score 40; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.89;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLPLQLILL 9
Db 6 LLPLQLILL 14
IIIIIIII
RESULT 2
ID IHA_TRIVU STANDARD; PRT; 361 AA.
AC Q7755;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Inhibin alpha chain precursor.
GN INHA.
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
OX NCBI_TaxID=9337;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99027340; PubMed=9801457;
RA Vanmontfort D., Fidler A.E., Heath D.A., Lawrence S.B., Tisdall D.J.,
RA Greenwood P.J., McNatty K.;
RA "cDNA sequence analysis, gene expression and protein localisation of
RT the inhibin alpha subunit of Australian brushtail possum (Trichosurus
RT vulpecula).";
RT J. Mol. Endocrinol. 21:141-152(1998).
RL J. Mol. Endocrinol. 21:141-152(1998).
CC -!- FUNCTION: INHIBINS AND ACTIVINS INHIBIT AND ACTIVATE,
CC RESPECTIVELY, THE SECRETION OF FOLLITROPIN BY THE PITUITARY GLAND.

```

```

CC INHIBINS/ACTIVINS ARE INVOLVED IN REGULATING A NUMBER OF DIVERSE
CC FUNCTIONS SUCH AS HYPOTHALAMIC AND PITUITARY HORMONE SECRETION,
CC GONADAL HORMONE SECRETION, GERM CELL DEVELOPMENT AND MATURATION,
CC EYTHROID DIFFERENTIATION, INSULIN SECRETION, NERVE CELL SURVIVAL,
CC EMBRYONIC AXIAL DEVELOPMENT OR BONE GROWTH, DEPENDING ON THEIR
CC SUBUNIT COMPOSITION. INHIBINS APPEAR TO OPPOSE THE FUNCTIONS OF
CC ACTIVINS.
CC -!- SUBUNIT: DIMERIC, LINKED BY ONE OR MORE DISULFIDE BONDS.
CC INHIBIN A IS A DIMER OF ALPHA AND BETA-A.
CC INHIBIN B IS A DIMER OF ALPHA AND BETA-B.
CC -!- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF033340; AAC63945.1; -.
CC InterPro: IPR002405; Inhibin_alpha.
CC InterPro: IPR001839; TGF-beta.
CC Pfam: PF00019; TGF-beta; 1.
CC PRINTS: PR00669; INHIBINA.
CC PRODom: PD000357; TGF-beta; 1.
CC SMART: SM00204; TGF-beta; 1.
CC PROSITE: PS00250; TGF_BETA_1; 1.
CC KW Growth factor; Hormone; Glycoprotein; Signal.
CC FT SIGNAL 1 21 POTENTIAL.
CC FT PROPEP 22 230 BY SIMILARITY.
CC FT CHAIN 231 361 INHIBIN ALPHA CHAIN.
CC FT DISULFID 260 323 BY SIMILARITY.
CC FT DISULFID 289 358 BY SIMILARITY.
CC FT DISULFID 293 360 BY SIMILARITY.
CC FT DISULFID 322 322 INTERCHAIN (BY SIMILARITY).
CC FT CARBOHYD 48 48 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).
CC FT CARBOHYD 266 266 N-LINKED (GLCNAC. .) (POTENTIAL).
CC SQ SEQUENCE 361 AA; 38945 MW; D66ICDF93CDAA87D CRC64;
Query Match 95.0%; Score 38; DB 1; Length 361;
Best Local Similarity 88.9%; Pred. No. 3.1;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLPLQLILL 9
Db 5 LLPLQLILL 13
IIIIIIII
RESULT 3
ID PSPD_BOVIN STANDARD; PRT; 369 AA.
AC P35246;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN SFTPD OR SFTPD.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.
RX TISSUE=Lung;
RX MEDLINE=93170856; PubMed=8436402;
RA Lim B.L., Lu J., Reid K.B.M.;
RA "Structural similarity between bovine conglutinin and bovine lung
RT surfactant protein D and demonstration of liver as a site of
RT synthesis of conglutinin.";
RL Immunology 78:159-165(1993).

```

CC -1- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
CC EXTENT OTHER ALPHA-GLUCOSYL MOETIES. IT COULD PARTICIPATE IN THE
CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
CC
CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC
CC -1- SUBCELLULAR LOCATION: Extracellular.
CC
CC -1- MISCELLANEOUS: PULMONARY SURFACTANT CONSISTS OF 90% LIPID AND 10%
CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
CC
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: X75911; CAA53510.1; --
CC PIR: S33603; S33603.
CC HSSP: P35247; 1B08.
CC InterPro: IPR000087; Collagen.
CC InterPro: IPR001304; lectin_c.
CC Pfam: PF01391; Collagen; 3.
CC Pfam: PF00059; lectin_c; 1.
CC SMART: SM00034; CLECT; 1.
CC PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
CC PROSITE: PS50041; C-TYPE-LECTIN_2; 1.
CC Glycoprotein: Calcium; Surface film; Gaseous exchange; Hydroxylation;
CC Signal; Lectin; Collagen; Repeat; Coiled coil.
CC SIGNAL 1 20
CC CHAIN 21 369
CC PULMONARY SURFACTANT-ASSOCIATED PROTEIN
CC D.
CC COLLAGEN-LIKE.
CC COILED COIL (POTENTIAL).
CC C-TYPE LECTIN (SHORT FORM).
CC BY SIMILARITY.
CC BY SIMILARITY.
CC BY SIMILARITY.
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC MOD_RES 90 90
CC MOD_RES 78 78
CC MOD_RES 87 87
CC MOD_RES 96 96
CC MOD_RES 99 99
CC MOD_RES 165 165
CC MOD_RES 171 171
CC MOD_RES 369 369
CC SEQUENCE 369 AA; 37361 MW; 07D88B24EOAEB2E3 CRC64;
CC
CC Query Match 85.0%; Score 34; DB 1; Length 369;
CC Best Local Similarity 77.8%; Pred. No. 19;
CC Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
CC
CC Qy 1 LLPLQLLLL 9
CC |||||:||||
CC Db 3 LLPLSVLLL 11
CC
CC RESULT 4
CC CONC_BOVIN CONG_BOVIN STANDARD; PRT; 371 AA.
CC ID CONG_BOVIN STANDARD; PRT; 371 AA.
CC AC P23805;
CC DT 01-NOV-1991 (Rel. 20, Created)
CC DT 01-FEB-1994 (Rel. 28, Last sequence update)
CC DT 01-OCT-1996 (Rel. 34, Last annotation update)
CC DE Conglutinin precursor.
CC GN CGN1.
CC OS Bos taurus (Bovine).
CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
CC OC Bovidae; Bovinae; Bos.
CC OX NCBI_TaxID=9913;

RN SEQUENCE FROM N.A.
RP MEDLINE-93213261; PubMed-8460993;
RX Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
RA "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-93277452; PubMed-7684896;
RA Lu J., Laursen S.B., Thiel S., Jensenius J.C., Reid B.M.;
RT "The cDNA cloning of conglutinin and identification of liver as a
RT primary site of synthesis of conglutinin in members of the Bovidae.";
RL Biochem. J. 292:157-162(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE-94215917; PubMed-8163202;
RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,
RA Tauber A.I., Sastry K.N.;
RT "Bovine conglutinin (BC) mRNA expressed in liver: cloning and
RT characterization of the BC cDNA reveals strong homology to surfactant
RT protein-D.";
RL Gene 141:277-281(1994).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Semen;
RX MEDLINE-94267222; PubMed-8207234;
RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,
RA Tauber A.I., Sastry K.N.;
RT "Bovine conglutinin gene exon structure reveals its evolutionary
RT relationship to surfactant protein-D.";
RL J. Immunol. 153:173-180(1994).
RN [5]
RP SEQUENCE OF 21-371.
RX MEDLINE-91131556; PubMed-1993651;
RA Lee Y.-M., Leiby K.R., Allar J., Paris K., Lerch B., Okarma T.B.;
RT "Primary structure of bovine conglutinin, a member of the C-type
RT animal lectin family.";
RL J. Biol. Chem. 266:2715-2723(1991).
RN [6]
RP PRELIMINARY SEQUENCE OF 21-52.
RX MEDLINE-87184551; PubMed-3566740;
RA Young N.M., Leon M.A.;
RT "The carbohydrate specificity of conglutinin and its homology to
RT proteins in the hepatic lectin family.";
RL Biochem. Biophys. Res. Commun. 143:645-651(1987).
CC -1- FUNCTION: CALCIUM-DEPENDENT LECTIN-LIKE PROTEIN WHICH BINDS TO A
CC YEAST CELL WALL EXTRACT & IMMUNE COMPLEXES THROUGH THE COMPLEMENT
CC COMPONENT (C3BI). IT IS CAPABLE OF BINDING NONREDUCING TERMINAL
CC N-ACETYLGLUCOSAMINE, MANNOSE, AND FUCOSE RESIDUES.
CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL: D14085; BAA03170.1; --
CC EMBL: X71774; CAA50665.1; --
CC EMBL: L18871; AAA20126.1; --
CC EMBL: U06860; AAB60624.1; --
CC EMBL: U06854; AAB60624.1; JOINED.
CC EMBL: U06855; AAB60624.1; JOINED.
CC EMBL: U06856; AAB60624.1; JOINED.
CC EMBL: U06857; AAB60624.1; JOINED.
CC EMBL: U06858; AAB60624.1; JOINED.
CC EMBL: U06859; AAB60624.1; JOINED.
CC PIR: A29416; A29416.

```

DR PIR; A23740; A23740.
DR PIR; JN0450; JN0450.
DR HSP; P35247; 1B08.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF01391; Collagen; 3.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN_1; 1.
DR PROSITE; PS00441; C-TYPE LECTIN_2; 1.
KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;
FT SIGNAL 1 20
FT CHAIN 21 371
FT DOMAIN 46 216
FT DOMAIN 273 371
FT MOD_RES 63 63
FT MOD_RES 87 87
FT MOD_RES 99 99
FT MOD_RES 135 135
FT MOD_RES 141 141
FT MOD_RES 159 159
FT MOD_RES 162 162
FT MOD_RES 198 198
FT MOD_RES 201 203
FT SITE 201 203
FT DISULFID 347 361
FT CARBOHYD 337 337
FT CONFLICT 173 173
FT CONFLICT 210 210
FT CONFLICT 218 218
FT CONFLICT 272 272
SQ SEQUENCE 371 AA; 37994 MW; 867BB41992544B1F CRC64;

Query Match 85.0%; Score 34; DB 1; Length 371;
Best Local Similarity 77.8%; Pred. No. 19;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
DB 3 LLPLSVLL 11

RESULT 5
TOXR_VIBCH STANDARD; PRT; 294 AA.
AC P15795; Q9KTB9;
DT 01-APR-1990 (Rel. 14, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Cholera toxin transcriptional activator.
GN TOXR OR VC0984.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CLASSICAL 569B / ATCC 25870 / SEROTYPE O1;
RX MEDLINE=87102874; PubMed=3802195;
RA Miller V.L., Taylor R.K., Mekalanos J.J.;
RA "Cholera toxin transcriptional activator toxR is a transmembrane DNA binding protein.";
RT Cell 48:271-279(1987).
RL [2]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
Fraser C.M.;
"DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae";
Nature 406:477-483(2000).
[3]
SEQUENCE OF 1-11 FROM N.A.
STRAIN=EL TOR E7946;
MEDLINE=91088618; PubMed=2124707;
Parsot C., Mekalanos J.J.;
"Expression of ToxR, the transcriptional activator of the virulence factors in Vibrio cholerae, is modulated by the heat shock response.";
Proc. Natl. Acad. Sci. U.S.A. 87:9898-9902(1990).
[4]
SEQUENCE OF 263-294 FROM N.A.
MEDLINE=91098651; PubMed=1898871;
Dirita V.J., Mekalanos J.J.;
"Periplasmic interaction between two membrane regulatory proteins, ToxR and ToxS, results in signal transduction and transcriptional activation.";
Cell 64:29-37(1991).
[5]
MUTAGENESIS.
RX MEDLINE=93015740; PubMed=1400230;
Ottemann K.M., Dirita V.J., Mekalanos J.J.;
"ToxR proteins with substitutions in residues conserved with OmpR fail to activate transcription from the cholera toxin promoter.";
J. Bacteriol. 174:6807-6814(1992).
CC -!- FUNCTION: THIS TRANSCRIPTION ACTIVATOR CONTROLS CHOLERA TOXIN, PILUS COLONIZATION FACTOR AND OUTER MEMBRANE PROTEIN EXPRESSION IN VIBRIO CHOLERA. IT BINDS TO THE 5'-TTTGTAT-3' TANDEMLY REPEATED DNA SEQUENCE IN CHOLERA TOXIN PROMOTER REGION. TOXS INTERACTS WITH THE C-TERMINAL PERIPLASMIC DOMAIN OF TOXR, STIMULATING ITS ACTIVITY. IT ACTIVATES TRANSCRIPTION AT THE PROMOTERS FOR TCP1 AND TCPA AND THIS IS PRESUMABLY VIA TOXT.
CC -!- SIMILARITY: TO E. COLI CADC, AND TO THE C-TERMINI OF A GROUP OF TRANSCRIPTIONAL ACTIVATORS (OMPR-LIKE PROTEINS).
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; M21249; AAA27549.1; -.
DR EMBL; AE004179; AAF94145.1; -.
DR EMBL; M58033; AAA27575.1; -.
DR EMBL; M62761; AAA63558.1; -.
DR PIR; A25970; A25970.
DR PIR; A38435; A38435.
DR TIGR; VC0984; -.
DR InterPro; IPR001867; Trans_reg_C.
DR Pfam; PF00486; trans_reg_C; 1.
KW Transcription regulation; DNA-binding; Activator; Transmembrane;
KW Trans-acting factor; Complete proteome.
FT DOMAIN 1 182 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 183 198
FT DOMAIN 199 294 PERIPLASMIC (POTENTIAL).
FT CONFLICT 134 134 S -> N (IN REF. 1).
FT CONFLICT 172 172 T -> A (IN REF. 1).
FT CONFLICT 184 184 L -> F (IN REF. 1).
FT CONFLICT 204 204 T -> S (IN REF. 1).
SQ SEQUENCE 294 AA; 32506 MW; 15EF85EA653C373E CRC64;

Query Match 82.5%; Score 33; DB 1; Length 294;
Best Local Similarity 77.8%; Pred. No. 24;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
DB 190 LLPLAVLL 198

```

```
RESULT 6
YA06_HUMAN
ID YA06_HUMAN STANDARD; PRT; 435 AA.
AC O60813;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical protein DJ845024.5 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Grafham D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Rhodes S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MAPE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AL022101; CAAL7880.1; -
DR EMBL; AL049680; CAB41252.1; -
KW Hypothetical protein.
FT NON_TER 435 435
SQ SEQUENCE 435 AA; 50539 MW; 3D19443032BBB494 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 435;
Best Local Similarity 87.5%; Pred. No. 36;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 8
DB 298 LVPLQILL 305
I:|||||

RESULT 7
YA04_HUMAN
ID YA04_HUMAN STANDARD; PRT; 500 AA.
AC O60810;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical protein DJ845024.2 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Grafham D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Rhodes S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MAPE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
```

```
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AL022101; CAAL7877.1; -
DR EMBL; AL049681; CAB41253.1; -
KW Hypothetical protein.
FT NON_TER 1
FT NON_TER 500 500
SQ SEQUENCE 500 AA; 58247 MW; 1895CD8A8F14B7C3 CRC64;

Query Match 82.5%; Score 33; DB 1; Length 500;
Best Local Similarity 87.5%; Pred. No. 41;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 8
DB 363 LVPLQILL 370
I:|||||

RESULT 8
APPL_MOUSE
ID APPL_MOUSE STANDARD; PRT; 653 AA.
AC Q03157;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-JUN-1994 (Rel. 29, Last annotation update)
DE Amyloid-like protein 1 precursor (APLP).
GN APLP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=93066322; PubMed=1279693;
RA Wasco W., Bupp K., Magendanz M., Gusella J.F., Tanzi R.E.,
RA Solomon F.;
RT Identification of a mouse brain cDNA that encodes a protein related
RT to the Alzheimer disease-associated amyloid beta protein precursor.*;
RL Proc. Natl. Acad. Sci. U.S.A. 89:10758-10762(1992).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. MAY BE PROCESSED
CC IN THE GOLGI COMPLEX; HOWEVER, IT IS NOT YET CLEAR WHETHER APLP
CC IS SECRETED.
CC -!- SIMILARITY: BELONGS TO THE APP FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; L04538; AAA37247.1; -
DR PIR; A46362; A46362.
DR HSPP; P05067; LMWP.
DR MGD; MGI:88046; Aipl1.
DR InterPro; IPR001868; A4_APP.
DR Pfam; PF02177; A4_EXTRA; 1.
DR PRINTS; PR00203; AMYLOIDA4.
DR SMART; SM00006; A4_EXTRA; 1.
DR PROSITE; PS00319; A4_EXTRA; 1.
DR PROSITE; PS00320; A4_INTRA; 1.
DR Glycoprotein; Transmembrane; Signal.
FT SIGNAL 1 20
FT DOMAIN 21 553
FT DOMAIN 21 583
FT TRANSMEM 584 606
FT DOMAIN 607 653
FT DOMAIN 643 646
FT DOMAIN 263 271
FT POLY-GLU.
```

FT CARBOHYD 464 464 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 554 554 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 653 AA; 72751 MW; 56516DC3EA40E4B0 CRC64;
Query Match 82.5%; Score 33; DB 1; Length 653;
Best Local Similarity 77.8%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLPLQILL 9
DB 22 LLPLSLLL 30
RESULT 9
MCTI_PAPHA STANDARD; PRT; 247 AA.
AC P52195;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Chymase precursor (EC 3.4.21.39) (Mast cell protease 1).
GN CMA1 OR CHM.
OS Papio hamadryas (Hamadryas baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Papio.
OX NCBI_TaxID=9557;
RN [1]
RP SEQUENCE FROM N.A., Husain A.;
RA Liao Y., Karnik S., to the EMBL/GenBank/DBJ databases.
RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAJOR SECRETED PROTEASE OF MAST CELLS WITH SUSPECTED
CC ROLES IN VASOACTIVE PEPTIDE GENERATION, EXTRACELLULAR MATRIX
CC DEGRADATION, AND REGULATION OF GLAND SECRETION.
CC -!- CATALYTIC ACTIVITY: SPECIFICITY SIMILAR TO CHYMOTRYPSIN.
CC -!- SUBCELLULAR LOCATION: MAST CELL GRANULES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. GRANZYME SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U38521; AAA91160.1; -;
DR EMBL; U38463; AAA91159.1; -;
DR HSSP; P23946; 1KLT.
DR MEROPS; S01.140; -;
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; TRYPSIN_DOM; 1.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 19
FT PROPEP 20 21 ACTIVATION PEPTIDE.
FT CHAIN 22 247
FT ACT_SITE 66 66 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 110 110 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 203 203 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 51 67 BY SIMILARITY.
FT DISULFID 144 209 BY SIMILARITY.
FT DISULFID 175 188 BY SIMILARITY.
FT CARBOHYD 80 80 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 103 103 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 247 AA; 27339 MW; E0EC15E0FA72FD8B CRC64;
Query Match 82.5%; Score 33; DB 1; Length 653;
Best Local Similarity 77.8%; Pred. No. 54;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Query Match 80.0%; Score 32; DB 1; Length 247;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLPLQILL 9
DB 3 LLPLPLLL 11
RESULT 10
LPSB_PERAM STANDARD; PRT; 256 AA.
AC P26305;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemolymph lipopolysaccharide-binding protein precursor (LPS-binding
DE protein) (LPS-BP).
OS Periplaneta americana (American cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blattodea; Blattidae; Periplaneta.
OX NCBI_TaxID=6978;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX TISSUE=Hemolymph;
RC MEDLINE=91302364; PubMed=1712779;
RA Jonori T., Natori S.;
RT "Molecular cloning of cDNA for lipopolysaccharide-binding protein
RT from the hemolymph of the American cockroach, Periplaneta americana.
RT Similarity of the protein with animal lectins and its acute phase
RT expression.";
RL J. Biol. Chem. 266:13318-13323(1991).
CC -!- FUNCTION: PARTICIPATES PROBABLY IN THE ELIMINATION OF FOREIGN
CC SUBSTANCES INVADING THE INSECT ABDOMINAL CAVITY, AND IN TRAPPING
CC INTRACELLULAR SYMBIOTS, WHEN THEY LEAK FROM THE MYCETOMES INTO
CC THE HEMOLYMPH.
CC -!- SUBCELLULAR LOCATION: SECRETED IN THE HEMOLYMPH.
CC -!- INDUCTION: BY INTRODUCTION OF FOREIGN CELLS INTO THE ABDOMINAL
CC CAVITY OF ADULT P.AMERICANA.
CC -!- MISCELLANEOUS: CALCIUM IS REQUIRED FOR LIPOPOLYSACCHARIDE BINDING.
CC -!- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D00711; BAA00616.1; -;
DR PIR; JQ0708; JQ0708.
DR HSSP; P35247; 1B08.
DR InterPro; IPR001304; lectin_c.
DR Pfam; PF00059; lectin_c; 1.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE_LECTIN_1; 1.
DR PROSITE; PS50041; C-TYPE_LECTIN_2; 1.
KW Glycoprotein; Lectin; Signal; Calcium.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 33
FT CHAIN 34 256 HEMOLYMPH LIPOPOLYSACCHARIDE-BINDING
FT CARBOHYD 56 56 N-LINKED (GLCNAC...) (PROBABLE).
FT DOMAIN 146 256 C-TYPE LECTIN (SHORT FORM).
FT DISULFID 148 252 BY SIMILARITY.
FT DISULFID 230 244 BY SIMILARITY.
SQ SEQUENCE 256 AA; 28420 MW; E8DF74F1AE639858 CRC64;
Query Match 80.0%; Score 32; DB 1; Length 256;
Best Local Similarity 66.7%; Pred. No. 32;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

```

QY      1  LLPLQILL 9
      |||| :||:
Db      7  LLPLSVLLM 15

RESULT 11
ARAJ_ECOLI
ID  ARAJ_ECOLI      STANDARD;      PRT;      394 AA.
AC  P23910;
DT  01-MAR-1992 (Rel. 21, Created)
DT  01-MAR-1992 (Rel. 21, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  Protein araj precursor.
GN  ARAJ OR B0396.
OS  Escherichia coli.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.
OX  NCBI_TaxID=562;
RN  [1]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=92078081; PubMed=1744033;
RA  Reeder T., Schleif R.;
RT  "Mapping, sequence, and apparent lack of function of araj, a gene of
RL  the Escherichia coli arabinose regulon.";
RN  J. Bacteriol. 173:7765-7771(1991).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12 / MG1655;
RX  MEDLINE=97426617; PubMed=9278503;
RA  Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA  Mau B., Shao Y.;
RT  "The complete genome sequence of Escherichia coli K-12.";
RL  Science 277:1453-1474(1997).
RN  [3]
RP  SEQUENCE FROM N.A.
RA  Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA  Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,
RA  Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RL  Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC  -1- FUNCTION: MAY BE INVOLVED IN EITHER THE TRANSPORT OR PROCESSING
CC  OF ARABINOSE POLYMERS.
CC  -1- INDUCTION: BY ARABINOSE.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  EMBL; M64787; AAA23474.1; -.
CC  EMBL; AE000145; AAC73499.1; -.
CC  EMBL; U73857; AAB18120.1; ALT_INIT.
CC  PIR; S27549; S27549.
CC  PIR; B43750; B43750.
CC  EcoGene; EG10060; araj.
CC  InterPro; IPR003662; sub.transporter.
CC  Pfam; PF000083; sugar.tr.1.
CC  Signal; Complete proteome.
CC  SIGNAL 19 POTENTIAL.
CC  FT CHAIN 20 394 PROTEIN ARAJ.
CC  SQ SEQUENCE 394 AA; 41926 MW; 40C1B0F1451A817A CRC64;

Query Match 80.0%; Score 32; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 50;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      3  PLQILL 9
      |||||
Db      309  PLQILL 315

RESULT 12
SEPL_HUMAN
ID  SEPL_HUMAN      STANDARD;      PRT;      412 AA.
AC  Q14242; Q12775;
DT  01-NOV-1997 (Rel. 35, Created)
DT  01-NOV-1997 (Rel. 35, Last sequence update)
DT  16-OCT-2001 (Rel. 40, Last annotation update)
DE  P-selectin glycoprotein ligand 1 precursor (PSGL-1) (Selectin P
DE  ligand) (CD162 antigen).
GN  SELPLG.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Placenta;
RX  MEDLINE=95332364; PubMed=7541799;
RA  Veldman G.M., Bean K.M., Cumming D.A., Eddy R.L., Sait S.N.J.,
RA  Shows T.B.;
RT  "Genomic organization and chromosomal localization of the gene
RL  encoding human P-selectin glycoprotein ligand.";
RN  J. Biol. Chem. 270:16470-16475(1995).
RN  [2]
RP  SEQUENCE FROM N.A.
RX  MEDLINE=94084793; PubMed=7505206;
RA  Sako D., Chang X.J., Barone K.M., Vachino G., White H.M., Shaw G.,
RA  Veldman G.M., Bean K.M., Ahern T.J., Furie B., Cumming D.A.,
RA  Larsen G.R.;
RT  "Expression cloning of a functional glycoprotein ligand for
RL  P-selectin.";
RN  Cell 75:1179-1186(1993).
RN  [3]
RP  SULFATION.
RX  MEDLINE=96028106; PubMed=7585949;
RA  Sako D., Comess K.M., Barone K.M., Camphausen R.T., Cumming D.A.,
RA  Shaw G.D.;
RT  "A sulfated peptide segment at the amino terminus of PSGL-1 is
RT  critical for P-selectin binding.";
RN  Cell 83:323-331(1995).
RN  [4]
RP  SULFATION.
RX  MEDLINE=96028107; PubMed=7585950;
RA  Pouyani T., Seed B.;
RT  "PSGL-1 recognition of P-selectin is controlled by a tyrosine
RT  sulfation consensus at the PSGL-1 amino terminus.";
RN  Cell 83:333-343(1995).
RN  [5]
RP  SULFATION.
RX  MEDLINE=96032682; PubMed=7559387;
RA  Wilkins P.P., Moore K.L., McEver R.P., Cummings R.D.;
RT  "Tyrosine sulfation of P-selectin glycoprotein ligand-1 is required
RT  for high affinity binding to P-selectin.";
RN  J. Biol. Chem. 270:22677-22680(1995).
CC  -1- FUNCTION: BINDS TO P-, E- AND L-SELECTINS. THE CALCIUM-DEPENDENT
CC  HIGH AFFINITY INTERACTION WITH P-SELECTIN MEDIATES THE TETHERING
CC  AND ROLLING OF NEUTROPHILS AND T-LYMPHOCYTES ON ENDOTHELIAL CELLS.
CC  -1- SUBUNIT: HOMODIMER; DISULFIDE-LINKED.
CC  -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC  -1- TISSUE SPECIFICITY: EXPRESSED ON NEUTROPHILS, MONOCYTES AND MOST
CC  LYMPHOCYTES.
CC  -1- PTM: HEAVILY O-GLYCOSYLATED. ALSO N-GLYCOSYLATED.
CC  -1- PTM: SULFATED IN THE N-TERMINAL REGION; SULFATION IS NECESSARY
CC  FOR P-SELECTIN BINDING.
CC  -1- DATABASE: NAME=PRO; NOTE=CD guide CD162 entry;
CC  WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd162.htm".
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U25956; AAA74577.1; -.
EMBL; U02297; AAC50061.1; -.
GlycoSuiteDB: Q14242; -.
MW; 600738; -.
Cell adhesion; Glycoprotein; Transmembrane; Signal; Repeat;
Polymorphism; Sulfation.
FT SIGNAL 1 17 POTENTIAL.
FT PROPEP 18 41
FT CHAIN 42 412 P-SELECTIN GLYCOPROTEIN LIGAND 1.
FT DOMAIN 18 320 EXTRACELLULAR (POTENTIAL).
FT TRANSNEM 321 341 POTENTIAL.
FT DOMAIN 342 412 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 65 65 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 111 111 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 302 302 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DOMAIN 122 261 12 X 10 AA TANDEM REPEATS.
FT REPEAT 122 131 1.
FT REPEAT 132 141 2.
FT REPEAT 142 151 3.
FT REPEAT 162 171 4.
FT REPEAT 182 191 5.
FT REPEAT 192 201 6.
FT REPEAT 202 211 7.
FT REPEAT 212 221 8.
FT REPEAT 222 231 9.
FT REPEAT 232 241 10.
FT REPEAT 242 251 11.
FT REPEAT 252 261 12.
FT MOD_RES 46 46 SULFATION (POTENTIAL).
FT MOD_RES 48 48 SULFATION (POTENTIAL).
FT MOD_RES 51 51 SULFATION (POTENTIAL).
FT VARIANT 132 141 MISSING (IN SHORT FORM; NOT AN ALTERNATIVE SPLICING).
FT FTIG-VAR 005611.
SQ SEQUENCE 412 AA; 43201 MW; A92A2A902DC9963A CRC64;

Query Match 80.0%; Score 32; DB 1; Length 412;
Best Local Similarity 75.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 LPLQLLLL 9
:|||||
Db 1 MPLQLLLL 8

RESULT 13
MM11_MOUSE STANDARD; PRT; 492 AA.
AC Q02853;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Stromelysin-3 precursor (EC 3.4.24.-) (Matrix metalloproteinase-11)
DE (MMP-11) (ST3) (SL-3).
GN MMP11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93054930; PubMed=1429845;
RA Lefebvre O., Wolf C., Limacher J.M., Hutin P., Wendling C.,
RA Lemer M., Basset P., Rio M.C.;
RT "The breast cancer-associated Stromelysin-3 gene is expressed during
mouse mammary gland apoptosis";
RL J. Cell Biol. 119:997-1002(1992).
RN [2]

RP REVISIONS.
RA Lefebvre O.;
RL Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases.
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 102-265.
RX PubMed=11254383;
RA Gall A.-L., Ruff M., Kannan R., Cunliffe P., Yiotakis A., Dive V.,
RA Rio M.-C., Basset P., Moras D.;
RT "Crystal structure of the Stromelysin-3 (MMP-11) catalytic domain
complexed with a phosphinic inhibitor mimicking the
transition-state";
RT J. Mol. Biol. 307:577-586(2001).
CC -!- FUNCTION: MAY PLAY AN IMPORTANT ROLE IN THE PROGRESSION OF
EPITHELIAL MALIGNANCIES.
CC -!- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY.
CC -!- TISSUE SPECIFICITY: SPECIFICALLY EXPRESSED IN THE MAMMARY GLAND
DURING APOPTOSIS.
CC -!- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
METALLOPROTEASE) ALSO KNOWN AS MATRIXIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL; Z12604; CAA78248.1; -.
DR PIR; A44399; A44399.
DR PDB; 1HV5; 28-MAR-01.
DR MEROPS; M10.007; -.
DR MGD; MGI:97008; Mmp11.
DR InterPro; IPR000585; Hemoexin.
DR InterPro; IPR001818; Matrixin.
DR InterPro; IPR000130; Zn_MTPeptidse.
DR Pfam; PF00045; hemoexin; 4.
DR Pfam; PF00413; Peptidase_M10; 1.
DR PRINTS; PR00138; MATRIXIN.
DR SMART; SM00120; HX; 4.
DR SMART; SM00235; ZnMc; 1.
DR PROSITE; PS00024; HEMOPEXIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; FALSE NEG.
KW Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium;
KW Collagen degradation; Extracellular matrix; Signal; 3D-structure.
FT SIGNAL 1 35 POTENTIAL.
FT PROPEP 36 101 ACTIVATION PEPTIDE (BY SIMILARITY).
FT CHAIN 102 492 STROMELYSIN-3.
FT DOMAIN 295 487 HEMOPEXIN-LIKE.
FT SITE 84 84 CYSTEINE SWITCH (BY SIMILARITY).
FT METAL 219 219 ZINC (CATALYTIC).
FT ACT_SITE 220 220 ZINC (CATALYTIC).
FT METAL 223 223 ZINC (CATALYTIC).
FT METAL 229 229 ZINC (CATALYTIC).
FT DISULFID 298 484 BY SIMILARITY.
SQ SEQUENCE 492 AA; 55441 MW; B54E260E4AB3D7C3 CRC64;
Query Match 80.0%; Score 32; DB 1; Length 492;
Best Local Similarity 77.8%; Pred. No. 63;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 LLPLQLLLL 9
|||||
Db 16 LLPLPLLLL 24

RESULT 14
MM17_HUMAN STANDARD; PRT; 606 AA.
ID MM17_HUMAN Q9ULZ9; Q14850;
DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Matrix metalloproteinase-17 precursor (EC 3.4.24.-) (MMP-17)
DE (Membrane-type matrix metalloproteinase 4) (MT-MMP 4) (Membrane-type-4
DE matrix metalloproteinase) (MT4-MMP).
GN MMP17 OR MT4MMP.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A. (LONG ISOFORM).
RP TISSUE=Monocytic Leukemia;
RC MEDLINE=96234364; PubMed=10471807;
RA Kajita M., Kinoh H., Ito N., Takamura A., Itoh Y., Okada A., Sato H.,
RA Seiki M.;
RT "Human membrane type-4 matrix metalloproteinase (MT4-MMP) is encoded
RT by a novel major transcript: isolation of complementary DNA clones
RT for human and mouse mt4-mmp transcripts.";
RL FEBS Lett. 457:353-356(1999).
RN [2]
RN SEQUENCE FROM N.A. (SHORT ISOFORM).
RP TISSUE=Breast carcinoma;
RC MEDLINE=96234364; PubMed=8640782;
RA Puentes X.S., Pendas A.M., Llano E., Velasco G., Lopez-Otin C.;
RA "Molecular cloning of a novel membrane-type matrix metalloproteinase
RT from a human breast carcinoma.";
RL Cancer Res. 56:944-949(1996).
RN [3]
RN SEQUENCE OF 129-302 FROM N.A., AND CHARACTERIZATION.
RP MEDLINE=20020281; PubMed=10551873;
RA Wang Y., Johnson A.R., Ye Q.-Z., Dyer R.D.;
RA "Catalytic activities and substrate specificity of the human membrane
RT type 4 matrix metalloproteinase catalytic domain.";
RL J. Biol. Chem. 274:33043-33049(1999).
RN [4]
RN GPI-ANCHOR.
RP MEDLINE=20036570; PubMed=10567400;
RA Itoh Y., Kajita M., Kinoh H., Mori H., Okada A., Seiki M.;
RT "Membrane type 4 matrix metalloproteinase (MT4-MMP, MMP-17) is a
RT glycosylphosphatidylinositol-anchored proteinase.";
RL J. Biol. Chem. 274:34260-34266(1999).
RN [5]
RN CHARACTERIZATION.
RP MEDLINE=20008793; PubMed=10543448;
RA Kolkenbrock H., Essers L., Ulbrich N., Will H.;
RT "Biochemical characterization of the catalytic domain of membrane-type
RT 4 matrix metalloproteinase.";
RL Biol. Chem. 380:1103-1108(1999).
CC [1]- FUNCTION: ENDOPEPTIDASE THAT DEGRADES VARIOUS COMPONENTS OF THE
CC EXTRACELLULAR MATRIX, SUCH AS FIBRIN. MAY BE INVOLVED IN THE
CC ACTIVATION OF MEMBRANE-BOUND PRECURSORS OF GROWTH FACTORS OR
CC INFLAMMATORY MEDIATORS, SUCH AS TUMOR NECROSIS FACTOR-ALPHA. MAY
CC ALSO BE INVOLVED IN TUMORAL PROCESS. NOT OBVIOUS IF ABLE TO
CC PROTEOLYTICALLY ACTIVATE PROGELATINASE A. DOES NOT HYDROLYSE
CC COLLAGEN TYPES I, II, III, IV AND V. GELATIN, FIBRONECTIN,
CC LAMININ, DECORIN NOR ALPHAI-ANTITRYPSIN.
CC [1]- CATALYTIC ACTIVITY: CLEAVES PRO-TNF-ALPHA AT THE 74-ALA-I-GLN-75
CC SITE.
CC [1]- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY (BY SIMILARITY).
CC [1]- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC [1]- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT/PENTE-TYPE FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC [1]- TISSUE SPECIFICITY: EXPRESSED IN BRAIN, LEUKOCYTES, COLON, OVARY
CC TESTIS AND BREAST CANCER. EXPRESSED ALSO IN MANY TRANSFORMED AND
CC NON-TRANSFORMED CELL TYPES.
CC [1]- PTM: THE PRECURSOR IS CLEAVED BY A FURIN ENDOPEPTIDASE (BY
CC SIMILARITY).
CC [1]- SIMILARITY: CONTAINS 1 HEMOXEPIN-LIKE DOMAIN.
CC [1]- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
CC METALLOPROTEASE) ALSO KNOWN AS MATRIXIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB021225; BAA82707.1; -;
DR EMBL: X89576; CAA61753.1; -;
DR HSSP: P08254; IUSN.
DR MEROPS: M10-017; -;
DR MIM: 602285; -;
DR InterPro: IPR000585; Hemopexin.
DR InterPro: IPR001818; Matrixin.
DR InterPro: IPR000130; Zn_MTPetdse.
DR Pfam: PF00045; hemopexin; 4.
DR Pfam: PF00413; Peptidase_M10; 1.
DR PRINTS: PR00138; MATRIXIN.
DR SMART: SM00120; HX; 4.
DR SMART: SM00235; ZnMc; 1.
DR PROSITE: PS00142; ZINC_PROTEASE; 1.
DR PROSITE: PS00546; CYSTEINE_SWITCH; FALSE_NEG.
DR PROSITE: PS00024; HEMOXEPIN; FALSE_NEG.
KW Hydrolase; Metalloprotease; Zinc; Calcium; zymogen; Signal;
KW GPI-anchor; Extracellular matrix; Calcium; Alternative splicing.
FT SIGNAL 1 38
FT PROPEP 39 128
FT CHAIN 129 568
FT PROPEP 569 606
FT DOMAIN 332 529
FT SITE 113 113
FT METAL 251 251
FT ACT_SITE 252 252
FT METAL 255 255
FT METAL 261 261
FT DISULFID 335 526
FT LIPID 568 568
FT DOMAIN 125 128
FT CARBOHYD 140 140
FT CARBOHYD 321 321
FT VARSPLIC 1 86
FT CONFLICT 205 210
FT CONFLICT 224 224
FT CONFLICT 228 228
SQ SEQUENCE 606 AA; 67123 MW; F3BF2EF21918F44C CRC64;
Query Match 80.0%; Score 32; DB 1; Length 606;
Best Local Similarity 77.8%; Pred. No. 78;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLPLQILL 9
Db 23 LLPLPLLL 31
RESULT 15
ACE_HUMAN
ID ACE_HUMAN STANDARD; PRT; 1306 AA.
AC P12821;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Angiotensin-converting enzyme, somatic isoform precursor (EC 3.4.15.1)
DE (ACE) (Dipeptidyl carboxypeptidase 1) (Kininase II) (CD143 antigen).
GN DCP1 OR DCP OR ACE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=89071703; PubMed=2849100;
RA Soubrier F., Alhenc-Gelas F., Hubert C., Allegrini J., John M.,

RA Tregear G., Corbol P.;
 RT "Two putative active centers in human angiotensin I-converting enzyme
 revealed by molecular cloning.";
 RL Proc. Natl. Acad. Sci. U.S.A. 85:9386-9390(1988).
 RN [2]
 RP SEQUENCE FROM N.A., AND VARIANTS S-261; W-561 AND S-1286.
 RX MEDLINE=99231580; PubMed=10319862;
 RA Rieder M.J., Taylor S.L., Clark A.G., Nickerson D.A.;
 RT "Sequence variation in the human angiotensin converting enzyme.";
 RL Nat. Genet. 22:59-62(1999).
 RN [3]
 RP PARTIAL SEQUENCE OF 30-46.
 RC TISSUE=Lung;
 RX MEDLINE=90110025; PubMed=2558109;
 RA Takeuchi K., Shimizu T., Ohishi N., Seyama Y., Takaku F.,
 RT Yotsumoto H.;
 RT "Purification of human lung angiotensin-converting enzyme by high-
 performance liquid chromatography: properties and N-terminal amino
 acid sequence.";
 RL J. Biochem. 106:442-445(1989).
 RN [4]
 RP ZINC-BINDING.
 RX MEDLINE=91308093; PubMed=1649623;
 RA Ehlers M.R., Riordan J.F.;
 RT "Angiotensin-converting enzyme: zinc- and inhibitor-binding
 stoichiometries of the somatic and testis isozymes.";
 RL Biochemistry 30:7118-7126(1991).
 CC -1- FUNCTION: CONVERTS ANGIOTENSIN I TO ANGIOTENSIN II BY RELEASE OF
 THE TERMINAL HIS-LEU. THIS RESULTS IN AN INCREASE OF THE
 VASOCONSTRICTOR ACTIVITY OF ANGIOTENSIN.
 CC -1- CATALYTIC ACTIVITY: Release of a C-terminal dipeptide,
 oligopeptide-l-Xaa-Xbb, when Xaa is not Pro, and Xbb is neither
 Asp nor Glu. Converts angiotensin I to angiotensin II.
 CC -1- COFACTOR: BINDS 2 ZINC IONS.
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
 CC -1- ALTERNATIVE PRODUCTS: THE TESTICULAR ANGIOTENSIN-CONVERTING ENZYME
 IS TRANSCRIBED FROM THE SAME GENE AS THE SOMATIC ISOFORM, PROBABLY
 FROM AN ALTERNATIVE START SITE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M2.
 CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD143 entry;
 WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd143.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: J04144; AAA51684.1; -
 DR EMBL: AF118569; AAD28560.1; -
 DR PIR: A31759; A31759.
 DR PIR: PQ0004; PQ0004.
 DR MIM: 106180; -
 DR InterPro: IPR001548; Peptidase_M2.
 DR InterPro: IPR000130; Zn_Mtpeptdse.
 DR Pfam: PF01401; Peptidase_M2; 2.
 DR PRINTS: PR00791; PEPDPTASEA.
 DR ProDom: PD004184; Peptidase_M2; 2.
 DR PROSITE: PS00142; ZINC_PROTEASE; 2.
 KW Hydrolase; Metalloprotease; Carboxypeptidase; Zinc; Dipeptidase;
 KW Glycoprotein; Transmembrane; Repeat; Signal; Alternative splicing;
 KW Polymorphism.
 FT SIGNAL 1 29
 FT CHAIN 30 1306 ANGIOTENSIN-CONVERTING ENZYME, SOMATIC
 FT ISOFORM.
 FT DOMAIN 30 1259 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 1260 1276 POTENTIAL.
 FT DOMAIN 1277 1306 CYTOPLASMIC (POTENTIAL).
 FT REPEAT 227 593
 FT REPEAT 825 1181
 FT METAL 390 390 ZINC 1 (CATALYTIC) (BY SIMILARITY).

FT ACT_SITE 391 391 1 (BY SIMILARITY).
 FT METAL 394 394 ZINC 1 (CATALYTIC) (BY SIMILARITY).
 FT METAL 988 988 ZINC 2 (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 989 989 2 (BY SIMILARITY).
 FT METAL 992 992 ZINC 2 (CATALYTIC) (BY SIMILARITY).
 FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 54 54 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 74 74 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 111 111 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 146 146 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 160 160 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 272 272 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 445 445 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 509 509 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 677 677 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 695 695 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 714 714 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 760 760 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 942 942 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 1191 1191 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT VARIANT 261 261 A -> S.
 FT VARIANT 561 561 /FTid=VAR_011707.
 FT VARIANT 1286 1286 R -> W.
 FT VARIANT 1286 1286 /FTid=VAR_011708.
 FT CONFLICT 35 35 /FTid=VAR_011709.
 FT CONFLICT 42 42 Q -> E (IN REF. 2).
 FT CONFLICT 42 42 D -> R (IN REF. 2).
 SQ SEQUENCE 1306 AA; 149714 MW; 1B33BCA7301A26AA CRC64;
 Query Match 80.0%; Score 32; DB 1; Length 1306;
 Best Local Similarity 77.8%; Pred. No. 1.7e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLPLQLLL 9
 DB 13 LLPLQLLL 21
 RESULT 16
 CFTR_XENLA
 ID CFTR_XENLA STANDARD; PRT; 1485 AA.
 AC P26363;
 DT 01-MAY-1992 (Rel. 22, Created)
 DT 01-MAY-1992 (Rel. 22, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Cystic fibrosis transmembrane conductance regulator (CFTR) (CAMP-
 dependent chloride channel).
 GN ABC7 OR CFTR.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
 OC Xenopodinae; Xenopus.
 OC NCBI_TaxID=8355;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93244789; PubMed=1284470;
 RA Tucker S.J., Tannahill D., Higgins C.F.;
 RT "Identification and developmental expression of the Xenopus laevis
 cystic fibrosis transmembrane conductance regulator gene.";
 RL Hum. Mol. Genet. 1:77-82(1992).
 CC -1- FUNCTION: INVOLVED IN THE TRANSPORT OF CHLORIDE IONS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: BELONGS TO THE ABC TRANSPORTER FAMILY. MRP SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----

```
or send an email to license@isb-sib.ch).
-----
CC EMBL: X65256; CAA46348.1; -.
CC PIR: S23756; S23756.
CC HSP: P13566; INBD.
CC InterPro: IPR003593; AAA.
CC InterPro: IPR001140; ABC_transporter_tmemb.
CC InterPro: IPR003439; ABC_transporter.
CC InterPro: IPR001687; ATP_GTP_A.
CC Pfam: PF00664; ABC_membrane; 2.
CC Pfam: PF00005; ABC_cran; 2.
CC SMART: SM00382; AAA; 2.
CC PROSITE: PS00211; ABC_TRANSPORTER; 1.
CC ATP-binding; Transmembrane; Transport; Glycoprotein; Repeat;
KW Ionic channel; Phosphorylation.
FT TRANSMEM 82 104 1 (POTENTIAL).
FT TRANSMEM 119 139 2 (POTENTIAL).
FT TRANSMEM 196 216 3 (POTENTIAL).
FT TRANSMEM 222 242 4 (POTENTIAL).
FT TRANSMEM 309 329 5 (POTENTIAL).
FT TRANSMEM 332 351 6 (POTENTIAL).
FT NP_BIND 459 466 ATP (BY SIMILARITY).
FT TRANSMEM 862 882 7 (POTENTIAL).
FT TRANSMEM 917 937 8 (POTENTIAL).
FT TRANSMEM 996 1016 9 (POTENTIAL).
FT TRANSMEM 1019 1039 10 (POTENTIAL).
FT TRANSMEM 1108 1128 11 (POTENTIAL).
FT TRANSMEM 1134 1155 11 (POTENTIAL).
FT NP_BIND 1247 1254 ATP (BY SIMILARITY).
FT CARBOHYD 894 894 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 900 900 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 1485 AA; 168895 MW; 2859E9B7DB233B89 CRC64;

Query Match 80.0%; Score 32; DB 1; Length 1485;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
DB 204 IAPLOVILL 212

RESULT 17
EFA2_MOUSE STANDARD; PRT; 209 AA.
AC P2801; 076020;
DT 01-OCT-1996 (Rel. 34, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)
DE (LERK-6) (ELF-1) (CEK7-ligand) (CEK7-L).
GN EFNA2 OR EPLG6 OR LERK6 OR ELF1 OR EPL6.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SWISS WEBSTER; TISSUE=Brain;
RX MEDLINE=95007776; PubMed=7522971;
RA Cheng H.J., Flanagan J.G.;
RT "Identification and cloning of ELF-1, a developmentally expressed
RL ligand for the Mek4 and Sek receptor tyrosine kinases."
RN Cell 79:157-168(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95181289; PubMed=7876076;
RA Shao H., Lou L., Pandey A., Verderame M.F., Siever D.A., Dixit V.M.;
RT "cDNA cloning and characterization of a CEK7 receptor
RT protein-tyrosine kinase ligand that is identical to the ligand
RT (ELF-1) for the Mek-4 and Sek receptor protein-tyrosine kinases."
RL J. Biol. Chem. 270:3467-3470(1995).
RL [3]

-|- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND
CC EPHA5.
CC -|- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -|- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U14941; AAA53636.1; -.
CC EMBL: U14752; AAA68520.1; -.
CC MGD: MGI:102707; Efn2.
CC InterPro: IPR001799; Ephrin.
CC Pfam: PF00812; Ephrin; 1.
CC PRINTS: PR01347; EPHRIN.
CC ProDom: PD002533; Ephrin; 1.
CC PROSITE: PS01299; EPHRIN; 1.
CC Glycoprotein; GPI-anchor; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 209 EPHRIN-A2.
FT CARBOHYD 38 38 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 184 184 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 209 AA; 23586 MW; F1997545F25B9ABC CRC64;

Query Match 77.5%; Score 31; DB 1; Length 209;
Best Local Similarity 77.8%; Pred. No. 41;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
DB 8 LLPLLLLL 16

RESULT 18
EFA2_HUMAN STANDARD; PRT; 213 AA.
AC Q43921; O76020;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ephrin-A2 precursor (EPH-related receptor tyrosine kinase ligand 6)
DE (LERK-6) (HEK7-ligand) (HEK7-L).
GN EFNA2 OR EPLG6 OR LERK6.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98126446; PubMed=9465306;
RA Carretti D.P., Nelson N.;
RT "Characterization of the genes for mouse LERK-3/Ephrin-A3 (Epl3),
RT mouse LERK-4/Ephrin-A4 (Epl4), and human LERK-6/Ephrin-A2 (EPLG6):
RT conservation of intron/exon structure."
RL Genomics 47:131-135(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Lamerdin J.E., McCreedy P.M., Skowronski E., Adamson A.W.,
RA Burkhardt-Schultz K., Gordon L., Kyle A., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Gaines J., Danganan L., Poundstone P.,
RA Christensen M., Georgescu A., Avila J., Liu S., Attix C., Andreise T.,
RA Trankhelm M., Amico-Keller G., Coefield J., Duarte S., Lucas S.,
RA Bruce R., Thomas P., Quan G., Kronmiller B., Arellano A.,
RA Montgomery M., Ow D., Nolan M., Trong S., Kobayashi A., Olsen A.O.,
RA Carraro A.V.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
```

```
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99045414; PubMed=9826538;
RA Aasheim H.C., Pedetour F., Grosgeorge J., Logtenberg T.;
RT "Cloning, chromosomal mapping, and tissue expression of the gene
RL encoding the human Eph-family kinase ligand ephrin-A2.";
RT Biochem. Biophys. Res. Commun. 252:378-382(1998).
CC -1- SUBUNIT: BINDS TO THE RECEPTOR TYROSINE KINASES EPHA3, EPHA4 AND
CC EPHA5.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE EPHRIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U92896; AAC39577.1; -
DR EMBL; U92893; AAC39577.1; JOINED.
DR EMBL; U92894; AAC39577.1; JOINED.
DR EMBL; AC004258; AAC04896.1; -
DR EMBL; AJ007292; CAA07435.1; -
DR MIM; 602756; -
DR InterPro; IPR001799; Ephrin.
DR Pfam; PF00812; Ephrin; 1.
DR PRINTS; PR01347; EPHRIN.
DR ProDom; PD002533; Ephrin; 1.
DR PROSITE; PS01299; EPHRIN; 1.
DR Glycoprotein; GPI-anchor; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 213 EPHRIN-A2.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 6 6 R -> A (IN REF. 3).
FT CONFLICT 25 26 RA -> PP (IN REF. 3).
FT CONFLICT 29 30 AA -> RR (IN REF. 3).
SQ SEQUENCE 213 AA; 23878 MW; 33C9FBI8168B2D0 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 213;
Best Local Similarity 77.8%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQLILL 9
Db 8 LLPLQLILL 16

RESULT 19
PPA5_RAT
ID PPA5_RAT STANDARD; PRT; 327 AA.
AC P29288;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Tartrate-resistant acid phosphatase type 5 precursor (EC 3.1.3.2) (TR-
DE AP) (Tartrate-resistant acid ATPase) (TRAPase).
GN ACP5.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Bone;
RX MEDLINE=92105136; PubMed=1722212;
RA Ek-Rylander B., Bill P., Norgaard M., Nilsson S., Andersson G.;
RT "Cloning, sequence, and developmental expression of a type 5,
RT tartrate-resistant, acid phosphatase of rat bone.";
J. Biol. Chem. 266:24684-24689(1991).
[2]
RN SEQUENCE OF 23-42.
RX MEDLINE=91314226; PubMed=1830446;
RA Ek-Rylander B., Bergman T., Andersson G.;
RT "Characterization of a tartrate-resistant acid phosphatase (ATPase)
RT from rat bone: hydrodynamic properties and N-terminal amino acid
RT sequence.";
J. Bone Miner. Res. 6:365-373(1991).
CC -1- FUNCTION: MAY PLAY A ROLE IN THE PROCESS OF BONE RESORPTION. THE
CC OSTEOCLASTIC TRAP ACTS ON NUCLEOTIDE TRI- AND DIPHOSPHATES WITH
CC HIGHER AFFINITY, COMPARED WITH OTHER SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O -> an
CC alcohol + phosphate.
CC -1- COFACTOR: BINDS TWO IRON ATOMS.
CC -1- SUBUNIT: EXISTS EITHER AS MONOMER OR, AFTER PROTEOLYTIC
CC PROCESSING, AS A DIMER OF TWO CHAINS LINKED BY DISULFIDE BOND(S).
CC -1- SUBCELLULAR LOCATION: Lysosomal.
CC -1- TISSUE SPECIFICITY: CHARACTERISTIC CONSTITUENT OF OSTEOCLASTS AND
CC SOME MONONUCLEAR PREOSTEOCLASTS. PREFERENTIALLY EXPRESSED IN
CC SKELETAL TISSUES.
CC -1- SIMILARITY: TO OTHER TARTRATE-RESISTANT ACID PHOSPHATASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M76110; AAA42305.1; -
DR PIR; A41720; A41720.
DR InterPro; IPR003266; PA_phosphatase.
DR InterPro; IPR000934; Ser_thr_phosphatse.
DR Pfam; PF02227; PA_phosphatase; 1.
DR Hydrolase; Glycoprotein; Iron; Lysosome; Signal.
FT SIGNAL 1 22
FT CHAIN 23 327 TARTRATE-RESISTANT ACID PHOSPHATASE TYPE
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 149 149 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 327 AA; 36726 MW; 5121A66A635ED854 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 327;
Best Local Similarity 88.9%; Pred. No. 65;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQLILL 9
Db 7 LLPLQLILL 15

RESULT 20
IBP2_HUMAN
ID IBP2_HUMAN STANDARD; PRT; 328 AA.
AC P18065; Q14619;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin-like growth factor binding protein 2 precursor (IGFBP-2)
DE (IBP-2) (IGF-binding protein 2).
GN IGFBP2 OR BP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Retina;
RX MEDLINE=91293227; PubMed=1712312;
RA Agarwal N., Hsieh C.L., Sills D., Swaroop M., Desai B., Francke U.,
RA Swaroop A.;
```

RT "Sequence analysis, expression and chromosomal localization of a
RT gene, isolated from a subtracted human retina cDNA library, that
RT encodes an insulin-like growth factor binding protein (IGFBP2).";
RL Exp. Eye Res. 52:549-561(1991).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 40-77.
RX MEDLINE=90368661; PubMed=1697583;
RA Zapf J., Klefer M., Merryweather J., Masiaz F., Bauer D., Born W.,
RA Fischer J.A., Foresch E.R.;
RT "Isolation from adult human serum of four insulin-like growth factor
RT (IGF) binding proteins and molecular cloning of one of them that is
RT increased by IGF I administration and in extrapancreatic tumor
RT hypoglycemia.";
RL J. Biol. Chem. 265:14892-14898(1990).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal liver;
RX MEDLINE=90060007; PubMed=2479552;
RA Binkert C., Landwehr J., Mary J.L., Schwander J., Heinrich G.;
RT "Cloning, sequence analysis and expression of a cDNA encoding a novel
RT insulin-like growth factor binding protein (IGFBP-2).";
RL EMBO J. 8:2497-2502(1989).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=91248211; PubMed=1710112;
RA Ehrenborg E., Vilhelmsdottir S., Bajjalica S., Larsson C., Stern I.,
RA Koch J., Brondum-Nielsen K., Luthman H.;
RT "Structure and localization of the human insulin-like growth factor-
RT binding protein 2 gene.";
RL Biochem. Biophys. Res. Commun. 176:1250-1255(1991).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=92293159; PubMed=1376411;
RA Binkert C., Margot J.B., Landwehr J., Heinrich G., Schwander J.;
RT "Structure of the human insulin-like growth factor protein-2
RT gene.";
RL Mol. Endocrinol. 6:826-836(1992).
CC -|- FUNCTION: IGF-BINDING PROTEINS PROLONG THE HALF-LIFE OF THE IGFs
CC AND HAVE BEEN SHOWN TO EITHER INHIBIT OR STIMULATE THE GROWTH
CC PROMOTING EFFECTS OF THE IGFs ON CELL CULTURE. THEY ALTER THE
CC INTERACTION OF IGFs WITH THEIR CELL SURFACE RECEPTORS.
CC -|- SUBCELLULAR LOCATION: Secreted.
CC -|- MISCELLANEOUS: BINDS IGF-II MORE THAN IGF-I.
CC -|- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC -|- SIMILARITY: BELONGS TO THE INSULIN-LIKE GROWTH FACTOR BINDING
CC PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S37730; AAB22308.1; -
DR EMBL: S37712; AAB22308.1; JOINED.
DR EMBL: S37722; AAB22308.1; JOINED.
DR EMBL: S37726; AAB22308.1; JOINED.
DR EMBL: M35410; AAA03246.1; -
DR EMBL: X16302; CAA34373.1; -
DR EMBL: M69241; AAA36048.1; -
DR EMBL: M69237; AAA36048.1; JOINED.
DR EMBL: M69239; AAA36048.1; JOINED.
DR EMBL: M69240; AAA36048.1; JOINED.
DR EMBL: A09809; CAA00862.1; -
DR PIR: S05389; S05389.
DR PIR: A35803; A35803.
DR PIR: JQ1029; JQ1029.
DR PIR: A41927; A41927.
DR PIR: A53957; A53957.

DR HSSP: P24593; 1BOE.
DR MIM: 146731; -
DR InterPro: IPR000867; IGFBP.
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00219; IGFBP; 1.
DR Pfam: PF00086; thyroglobulin_1; 1.
DR SMART: SM00121; IB; 1.
DR SMART: SM00211; TV; 1.
DR PROSITE: PS00222; IGF-BINDING; 1.
DR PROSITE: PS00484; THYROGLOBULIN_1; 1.
KW Growth factor binding; Signal.
FT SIGNAL 1 39
FT CHAIN 40 328 INSULIN-LIKE GROWTH FACTOR BINDING
FT PROTEIN 2.
FT DOMAIN 260 309 THYROGLOBULIN TYPE I.
FT SITE 304 306 CELL ATTACHMENT SITE.
FT CONFLICT 60 60 P -> R (IN REF. 4).
FT CONFLICT 320 320 R -> C (IN REF. 3).
FT CONFLICT 323 323 H -> D (IN REF. 4).
SQ SEQUENCE 328 AA; 35137 MW; 4E6BDF6D805C8853 CRC64;
Query Match 77.5%; Score 31; DB 1; Length 328;
Best Local Similarity 77.8%; Pred. No. 65;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLPLQIILL 9
DB 20 LLPLLLLL 28
RESULT 21
YFJD_ECOLI STANDARD; PRT; 428 AA.
AC P37908; P76600; P76601; P77009;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfjd.
GN YFJD OR B2612/B2613.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=K12 / MGL655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [3]
RP SEQUENCE OF 322-428 FROM N.A.
RX STRAIN=B178;
RX MEDLINE=88319942; PubMed=3045760;
RA Lipinska B., King J., Ang D., Georgopoulos C.;
RT "Sequence analysis and transcriptional regulation of the Escherichia

```
RT col1 grpE gene, encoding a heat shock protein.";
```

[4]

RA IDENTIFICATION.

RP Rudd K.E.;

RA Unpublished observations (AUG-1994).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

CC -1- SIMILARITY: BELONGS TO THE UPF0053 FAMILY. STRONG, TO

CC H. INFLUENZAE HT10107.

CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.

CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A

CC FRAMESHIFT IN POSITION 199 THAT PRODUCES TWO ORFS.

CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO NUMEROUS

CC FRAMESHIFTS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AE000347; AAC75661.1; ALT_FRAME.

DR EMBL; AE000347; AAC75662.1; ALT_FRAME.

DR EMBL; D90888; BAA16497.1; ALT_INIT.

DR EMBL; X07863; -: NOT_ANNOTATED_CDS.

DR ECoGene; EGI2442; yfjd.

DR InterPro; IPR000644; CBS.

DR InterPro; IPR002550; DUF21.

DR Pfam; PF00571; CBS; 2.

DR Pfam; PF01595; DUF21; 1.

DR SMART; SM00116; CBS; 1.

DR Hypothetical protein; CBS domain; Repeat; Transmembrane;

KW Complete proteome.

KW TRANSMEM 4 24 POTENTIAL.

FT TRANSMEM 65 85 POTENTIAL.

FT TRANSMEM 92 112 POTENTIAL.

FT TRANSMEM 130 150 POTENTIAL.

SQ SEQUENCE 428 AA; 48044 MW; 38456865EDBCB151 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 428;

Best Local Similarity 66.7%; Pred. No. 86;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9

Db 132 LAPLQILMM 140

RESULT 22

NCBI_HUMAN STANDARD; PRT; 461 AA.

ID NCBI_HUMAN

AC Q02818; Q15838; Q9BURI;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Nucleobindin 1 precursor (CALNUC).

GN NUCB1 OR NUC.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92392352; PubMed=1520323;

RA Miura K., Titani K., Kurosawa Y., Kanai Y.;

RT "Molecular cloning of nucleobindin, a novel DNA-binding protein that

RT contains both a signal peptide and a leucine zipper structure.";

RL Biochem. Biophys. Res. Commun. 187:375-380(1992).

RN [2]

RP SEQUENCE FROM N.A., VARIANTS 13-S-DEL-14 AND ARG-399, AND REVISIONS

RP TO 2999-300.

TISSUE=Placenta;

MEDLINE=96299707; PubMed=8661046;

RA Miura K., Hirai M., Kanai Y., Kurosawa Y.;

RT "Organization of the human gene for nucleobindin (NUC) and its

RT chromosomal assignment to 19q13.2-q13.4.";

RL Genomics 34:181-186(1996).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE=Lung;

RA Strausberg R.;

RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.

RN [4]

RP CHARACTERIZATION.

RX MEDLINE=96026433; PubMed=7589456;

RA Mochizuki N., Hibi M., Kanai Y., Insel P.A.;

RT "Interaction of the protein nucleobindin with G alpha 12, as revealed

RT by the yeast two-hybrid system.";

RL FEBS Lett. 373:155-158(1995).

CC -1- FUNCTION: Major calcium-binding protein of the Golgi. May have a

CC role in calcium homeostasis (By similarity).

CC -1- SUBUNIT: Interacts with GNAI2.

CC -1- SUBCELLULAR LOCATION: In the cis-Golgi, tightly associated with

CC the luminal side of the membrane. A small fraction of the protein

CC may be cytoplasmic (By similarity).

CC -1- TISSUE SPECIFICITY: Expressed both in fetal and adult heart, lung,

CC liver, kidney and brain, and in adult skeletal muscle, placenta

CC and pancreas.

CC -1- MISCELLANEOUS: Discovered as DNA-binding protein in the serum of

CC lupus-prone mice.

CC -1- SIMILARITY: BELONGS TO THE NUCLEOBINDIN FAMILY.

CC -1- SIMILARITY: CONTAINS 2 EF-HAND CALCIUM-BINDING DOMAINS.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M96824; AAA36383.1; -.

DR EMBL; U31342; AAB60431.1; -.

DR EMBL; U31336; AAB60431.1; JOINED.

DR EMBL; U31337; AAB60431.1; JOINED.

DR EMBL; U31338; AAB60431.1; JOINED.

DR EMBL; U31340; AAB60431.1; JOINED.

DR EMBL; U31341; AAB60431.1; JOINED.

DR EMBL; BC002356; AAH02356.1; -.

DR MIM; 601323; -.

DR InterPro; IPR002048; EF-hand.

DR Pfam; PF00036; ehand; 2.

DR PROSITE; PS00018; EF_HAND; 2.

DR Calcium-binding; DNA-binding; Golgi stack; Repeat; Signal;

KW Polymorphism.

FT SIGNAL 1 26 POTENTIAL.

FT CHAIN 27 461 NUCLEOBINDIN 1.

FT DNA_BIND 172 218 POTENTIAL.

FT CA_BIND 253 264 EF-HAND 1 (POTENTIAL).

FT CA_BIND 305 316 EF-HAND 2 (POTENTIAL).

FT DOMAIN 228 321 BINDS GNAI2 (BY SIMILARITY).

FT DOMAIN 401 407 POLY-GLN.

FT VARIANT 13 14 LP -> S.

FT VARIANT 399 399 /FTID=VAR_012151.

FT VARIANT 399 399 Q -> R.

FT CONFLICT 299 300 /FTID=VAR_012152.

FT CONFLICT 385 385 HV -> QL (IN REF. 1).

FT CONFLICT 390 391 K -> Q (IN REF. 3).

FT CONFLICT 390 391 LL -> QQ (IN REF. 3).

SQ SEQUENCE 461 AA; 53821 MW; 01B7D04DF748A42A CRC64;

Query Match 77.5%; Score 31; DB 1; Length 461;

Best Local Similarity 77.8%; Pred. No. 92;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||
Db 12 LLPLLLLL 20

RESULT 23
HRA1_HUMAN STANDARD; PRT; 480 AA.
AC Q32743: O9UN5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Serine protease HTRA1 precursor (EC 3.4.21.-) (L56).
GN PRS11 OR HTRA1 OR HTRA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP
RC TISSUE=placenta;
RX MEDLINE=97131593; PubMed=8977104;
RA Zumbunn J., Trueb B.;
RT "Primary structure of a putative serine protease specific for IGF-
binding proteins.";
RL FEBS Lett. 398:187-192(1996).
[2]
SEQUENCE FROM N.A.
RA Crowl R.M., Luk D., Milnamow M.;
RT "Genomic organization and promoter characterization of the human HTRA
(PRS11) gene.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 144-480 FROM N.A., AND MUTAGENESIS OF SER-328.
RP
RC TISSUE=cartilage;
RX MEDLINE=99069438; PubMed=9852107;
RA Hu S.I., Carozza M., Klein M., Nantermet P., Luk D., Crowl R.M.;
RT "Human HTRA, an evolutionarily conserved serine protease identified as
a differentially expressed gene product in osteoarthritis cartilage.";
RL J. Biol. Chem. 273:34406-34412(1998).
CC -1- FUNCTION: Protease that regulate the availability of IGFs by
cleaving IGF-binding proteins.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: Expressed in a variety oftissues, with
strongest expression in placenta.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C; ALSO KNOWN AS THE
DEGP/DEGO/DEGS FAMILY.
CC -1- SIMILARITY: CONTAINS 1 IGFBP DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 KAZAL-LIKE DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC
CC EMBL: Y07921; CAA69226.1; -
CC EMBL: AF157623; AAD41525.1; -
CC EMBL: AF097709; AAC97211.1; -
CC HSP: P80424; IAN1.
CC MEROPS: S01.277; -
CC MIM: 602194; -
CC InterPro: IPR000867; IGFBP.
CC InterPro: IPR001478; PDZ.
CC InterPro: IPR001940; Protease2C.
CC InterPro: IPR001254; Trypsin.
CC InterPro: IPR002350; Kazal.
CC Pfam: PF00219; IGFBP; 1.
CC Pfam: PF00050; kazal; 1.

DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00121; IB; 1.
DR SMART: SM00280; KAZAL; 1.
DR SMART: SM00228; PDZ; 1.
DR PROSITE: PS0106; PDZ; 1.
KW Hydrolase; Serine protease; Growth factor binding; Signal.
FT SIGNAL 1 22
FT CHAIN 23 480
FT DOMAIN 37 94
FT DOMAIN 101 155
FT DOMAIN 204 364
FT DOMAIN 365 467
FT ACT_SITE 220 220
FT ACT_SITE 250 250
FT ACT_SITE 328 328
FT MUTAGEN 328 328
FT CONFLICT 323 323
SQ SEQUENCE 480 AA; 51286 MW; CA20A99480FE2330 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 480;
Best Local Similarity 77.8%; Pred. No. 96;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||
Db 8 LLPLLLLL 16

RESULT 24
PVR2_MOUSE STANDARD; PRT; 530 AA.
ID PVR2_MOUSE
AC P32507; Q62096;
DT 01-OCT-1993 (Rel. 27, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Poliovirus receptor related protein 2 precursor (Murine herpesvirus
entry protein B) (MHVSB) (Nectin 2) (Poliovirus receptor homolog).
GN PVR2 OR PVS OR PVR OR MPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
SEQUENCE FROM N.A. (ISOFORM ALPHA).
RX MEDLINE=92219365; PubMed=1560525;
RA Morrison M.E., Racaniello V.R.;
RT "Molecular cloning and expression of a murine homolog of the human
poliovirus receptor gene.";
RL J. Virol. 66:2807-2813(1992).
[2]
SEQUENCE FROM N.A. (ISOFORM BETA).
RX STRAIN=C57BL/6; TISSUE=Brain;
RX MEDLINE=94179228; PubMed=8132569;
RA Aoki J., Koike S., Ise I., Sato-Yoshida Y., Nomoto A.;
RT "Amino acid residues on human poliovirus receptor involved in
interaction with poliovirus.";
RL J. Biol. Chem. 269:8431-8438(1994).
[3]
CHARACTERIZATION.
RX MEDLINE=99214397; PubMed=10196354;
RA Shukla D., Rowe C.L., Dong Y., Racaniello V.R., Spear P.G.;
RT "The murine homolog (Mph) of human herpesvirus entry protein B (HvEB)
mediates entry of pseudorabies virus but not herpes simplex virus
types 1 and 2.";
RT J. Virol. 73:4493-4497(1999).
CC -1- FUNCTION: RECEPTOR FOR ALPHAHERPESVIRUS (SUCH AS MURINE HSV) ENTRY
INTO CELLS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA AND BETA (SHOWN HERE);
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: BRAIN, SPINAL CORD, SPLEEN, KIDNEY, HEART AND

DE Neurexin 2-beta precursor (Neurexin II-beta).
GN NRXN2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP CONCEPTUAL TRANSLATION OF 1-89.
RA Roehrbert B.;
RL Unpublished observations (OCT-2001).
RN [2]
RP SEQUENCE OF 90-666 FROM N.A.
RA Rowen L., Madan A., Qin S., Baradarani L., Birditt B., Bloom S.,
RA Burke J., Dors M., Fleetwood P., Kaur A., Madan A., Nesbitt R.,
RA Pate D., Hood L.;
RT "Sequencing of human neurexin II gene";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NEURONAL CELL SURFACE PROTEIN THAT MAY BE INVOLVED IN
CC CELL RECOGNITION AND CELL ADHESION.
CC -1- SUBUNIT: THE CYTOPLASMIC C-TERMINAL REGION BINDS TO CASK. SPECIFIC
CC ISOFORMS BIND ALPHA-DYSTROGLYCAN (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -1- ALTERNATIVE PRODUCTS: MULTIPLE ISOFORMS (BY SIMILARITY). THE ALPHA
CC (AC 09P2S2) AND BETA ISOFORMS DIFFER IN THEIR N-TERMINUS BY USE OF
CC ALTERNATIVE PROMOTERS IN THE NRXN2 GENE.
CC -1- DOMAIN: ALTERNATIVE SPLICING IN THE LAMININ G-LIKE DOMAIN
CC REGULATES BINDING TO ALPHA-DYSTROGLYCAN.
CC -1- SIMILARITY: CONTAINS 1 LAMININ G-LIKE DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NEUREXIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch).

DR EMBL: AC044790; -; NOT_ANNOTATED_CDS.
DR MIM: 600566;
DR InterPro: IPR001791; Laminin_G.
DR SMART: SM00294; 4.1m; 1.
DR SMART: SM00282; LamG; 1.
DR PROSITE: PS50025; LAM_G_DOMAIN; 1.
KW Signal: Transmembrane; Cell adhesion; Glycoprotein;
KW Alternative splicing.
FT SIGNAL 1 50 BY SIMILARITY.
FT CHAIN 51 666 NEUREXIN 2-BETA.
FT DOMAIN 51 590 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 591 611 POTENTIAL.
FT DOMAIN 612 666 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 91 299 LAMININ G-LIKE.
FT DOMAIN 322 325 POLY-THR.
FT DOMAIN 328 331 POLY-THR.
FT DOMAIN 398 401 POLY-PRO.
FT DOMAIN 598 601 POLY-ALA.
SQ SEQUENCE 666 AA; 70927 MW; FD0B2FD35F63C6C CRC64;

Query Match 77.5%; Score 31; DB 1; Length 666;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQLLLL 9
Db 36 LLPLLLLLL 44

RESULT 27
PTPE_HUMAN
ID PTPE_HUMAN STANDARD; PRT; 700 AA.
AC P23469;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protein-tyrosine phosphatase epsilon precursor (EC 3.1.3.48) (R-PTP-
DE epsilon).
GN PTPRE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=91006018; PubMed=2170109;
RA Krueger N.X., Streuli M., Saito H.;
RT "Structural diversity and evolution of human receptor-like protein
RT tyrosine phosphatases";
RL EMBL J. 9:3241-3252(1990).
CC -1- CATALYTIC ACTIVITY: Protein tyrosine phosphate + H(2)O -> protein
CC tyrosine + phosphate.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: CONTAINS 2 PROTEIN-TYROSINE PHOSPHATASE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@sib-sib.ch).

DR EMBL: X54134; CAA38069.1;
DR PIR: S12053; S12053.
DR HSP: P18052; 1YFO.
DR MIM: 600926;
DR InterPro: IPR000387; TYR_phosphatase.
DR InterPro: IPR000242; Tyr_prot_phptase.
DR Pfam: PF00102; Y_phosphatase; 2.
DR PRINTS: PR00700; PRTYPHPTASE.
DR SMART: SM00194; PTPC; 2.
DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 2.
DR PROSITE: PS00056; TYR_PHOSPHATASE_2; 2.
DR PROSITE: PS00055; TYR_PHOSPHATASE_PTP; 2.
KW Glycoprotein; Transmembrane; Hydrolase; Phosphorylation; Repeat;
KW Signal.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 700 PROTEIN-TYROSINE PHOSPHATASE EPSILON.
FT DOMAIN 20 46 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 47 69 POTENTIAL.
FT DOMAIN 70 700 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 154 393 PROTEIN-TYROSINE PHOSPHATASE 1.
FT DOMAIN 394 700 PROTEIN-TYROSINE PHOSPHATASE 2.
FT ACT_SITE 335 335 BY SIMILARITY.
FT ACT_SITE 630 630 BY SIMILARITY.
FT CARBOHYD 23 23 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 30 30 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 700 AA; 80641 MW; D096BACDEA65708 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 700;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQLLLL 9
Db 52 LLPLLLLLL 60

RESULT 28
CHC2_HUMAN
ID CHC2_HUMAN STANDARD; PRT; 1007 AA.
AC Q9Y5F4; Q9Y5F4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protocadherin alpha C2 precursor (PCDH-alpha-C2).

```
GN PCDHAC2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (SHORT AND LONG ISOFORMS).
RC TISSUE=Brain;
RX MEDLINE=99308636; PubMed=10380929;
RA Wu Q., Maniatis T.;
RT "A striking organization of a large family of human neural cadherin-
RL like cell adhesion genes.";
RL Cell 97:779-790(1999).
CC -|- FUNCTION: POTENTIAL CALCIUM-DEPENDENT CELL-ADHESION PROTEIN.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -|- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -|- SIMILARITY: CONTAINS 6 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF152304; AAD43698.1; -.
DR EMBL: AF152474; AAD43735.1; -.
DR MIM: 606321; -.
DR MIM: 604966; -.
DR HSSP: P15116; INCJ.
DR InterPro: IPR002126; Cadherin.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA; 6.
DR PROSITE: PS00232; CADHERIN_1; 5.
DR PROSITE: PS0268; CADHERIN_2; 6.
DR Calcium-binding; Cell adhesion; Glycoprotein; Signal; Repeat;
KW Transmembrane; Multigene family; Alternative splicing.
FT SIGNAL 1 42
FT CHAIN 43 1007
FT DOMAIN 43 708
FT TRANSMEM 709 729
FT POTENTIAL.
FT DOMAIN 730 1007
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 43 148
FT CADHERIN 1.
FT DOMAIN 149 257
FT DOMAIN 258 365
FT CADHERIN 2.
FT DOMAIN 374 469
FT CADHERIN 3.
FT DOMAIN 470 579
FT CADHERIN 4.
FT DOMAIN 594 691
FT CADHERIN 5.
FT DOMAIN 800 907
FT CADHERIN 6.
FT POLY-LYS.
FT DOMAIN 856 951
FT CADHERIN 7.
FT REPEAT 856 859
FT 4 X PXXP REPEATS.
FT PXXP 1.
FT REPEAT 889 892
FT PXXP 2.
FT REPEAT 930 933
FT PXXP 3.
FT REPEAT 948 951
FT PXXP 4.
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 280 280
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 436 436
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 586 586
FT N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 657 657
FT VARSPLIC 856 884
FT LQTHAFVTHPTSCDALLSH (IN SHORT ISOFORM).
FT VARSPLIC 885 1007
FT MISSING (IN SHORT ISOFORM).
FT SEQUENCE 1007 AA; 109449 MW; 39BF2356923E3B67 CRC64;
SQ
Query Match 77.5%; Score 31; DB 1; Length 1007;
Best Local Similarity 77.8%; Pred. No. 2.1e-02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLPQLILL 9
Db 25 LLPLLLLL 33
```

```
RESULT 29
PC16_HUMAN
ID PC16_HUMAN STANDARD; PRT; 3298 AA.
AC Q96JQ0; O15098;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Protocadherin 16 precursor (Cadherin 19) (Cadherin fibroblast 1).
GN PCDH16 OR CDH19 OR FIB1 OR KIAA1773.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=21481446; PubMed=11597768;
RA Nakajima D., Nakayama M., Kikuno R., Hirose M., Nagase T., Ohara O.;
RT "Identification of three novel non-classical cadherin genes through
RT comprehensive analysis of large cDNAs.";
RL Brain Res. Mol. Brain Res. 94:85-95(2001).
RN [2]
RP SEQUENCE OF 434-570 FROM N.A.
RX MEDLINE=97342499; PubMed=9199196;
RA Matsuyoshi N., Imamura S.;
RT "Multiple cadherins are expressed in human fibroblasts.";
RL Biochem. Biophys. Res. Commun. 235:355-358(1997).
CC -|- FUNCTION: Calcium-dependent cell-adhesion protein (Potential).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein (By similarity).
CC -|- TISSUE SPECIFICITY: Fibroblast specific.
CC -|- SIMILARITY: CONTAINS 27 CADHERIN DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AB053446; BAB61903.1; -.
DR EMBL: AB000895; BAA21133.1; -.
DR MIM: 603057; -.
DR HSSP: P15116; INCJ.
DR InterPro: IPR002126; Cadherin.
DR PRINTS: PF00028; cadherin; 1.
DR PRINTS: PR00205; CADHERIN.
DR SMART: SM00112; CA; 1.
DR PROSITE: PS00232; CADHERIN_1; 18.
DR PROSITE: PS0268; CADHERIN_2; 27.
DR Cell adhesion; Glycoprotein; Transmembrane; Calcium-binding; Repeat;
KW Signal; Multigene family.
FT SIGNAL 1 42
FT CHAIN 43 3298
FT POTENTIAL.
FT DOMAIN 43 2940
FT EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 2941 2961
FT DOMAIN 2962 3298
FT CYTOPLASMIC (POTENTIAL).
FT DOMAIN 43 143
FT CADHERIN 1.
FT DOMAIN 144 255
FT CADHERIN 2.
FT DOMAIN 256 362
FT CADHERIN 3.
FT DOMAIN 367 472
FT CADHERIN 4.
FT DOMAIN 474 578
FT CADHERIN 5.
FT DOMAIN 579 685
FT CADHERIN 6.
FT DOMAIN 686 790
FT CADHERIN 7.
FT DOMAIN 791 894
FT CADHERIN 8.
FT DOMAIN 895 1000
FT CADHERIN 9.
FT DOMAIN 1001 1111
FT CADHERIN 10.
FT DOMAIN 1112 1211
FT CADHERIN 11.
FT DOMAIN 1218 1324
FT CADHERIN 12.
FT DOMAIN 1333 1436
FT CADHERIN 13.
FT DOMAIN 1437 1546
FT CADHERIN 14.
```

```
FT DOMAIN 1547 1649 CADHERIN 15.
FT DOMAIN 1650 1751 CADHERIN 16.
FT DOMAIN 1752 1855 CADHERIN 17.
FT DOMAIN 1856 1960 CADHERIN 18.
FT DOMAIN 1965 2068 CADHERIN 19.
FT DOMAIN 2069 2171 CADHERIN 20.
FT DOMAIN 2172 2277 CADHERIN 21.
FT DOMAIN 2278 2376 CADHERIN 22.
FT DOMAIN 2377 2482 CADHERIN 23.
FT DOMAIN 2483 2602 CADHERIN 24.
FT DOMAIN 2603 2706 CADHERIN 25.
FT DOMAIN 2707 2813 CADHERIN 26.
FT DOMAIN 2814 2933 CADHERIN 27.
FT CARBOHYD 217 217 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 256 256 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 402 402 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 584 584 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1249 1249 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1521 1521 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1718 1718 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1996 1996 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2361 2361 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2428 2428 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2569 2569 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2761 2761 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2792 2792 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2862 2862 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 3298 AA; 346175 MW; 6E8ED28BEF0795DB CRC64;

Query Match 77.5%; Score 31; DB 1; Length 3298;
Best Local Similarity 77.8%; Pred. No. 6.9e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPLQLILL 9
Db 22 LLPLLLLL 30

RESULT 30
POL_FLV STANDARD; PRT; 128 AA.
AC P10273;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Pol polypeptide [Contains: Protease (EC 3.4.23.-); Reverse
DE transcriptase (EC 2.7.7.49); Endonuclease] (Fragment).
GN POL.
OS Feline leukemia virus.
OC Viruses; Retroviral viruses; Retroviridae; Gammaretrovirus.
OX NCBI_TaxID=11768;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84216470; PubMed=6328019;
RA Laprevotte I., Hampe A., Sherr C.J., Galibert F.;
RT "Nucleotide sequence of the gag gene and gag-pol junction of feline
RT leukemia virus.";
RL J. Virol. 50:884-894(1984).
CC -!- PTM: SPECIFIC ENZYMATIC CLEAVAGES IN VIVO YIELD MATURE PROTEINS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A2; ALSO KNOWN AS THE
CC RETROPEPSIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; K01803; AAA43056.2; -.
CC HSSP; P03369; 3HVP.
CC MEROPS; A02.008; -.
CC -----
```

```
DR InterPro; IPR001995; Asp_prot_retrov.
DR InterPro; IPR001969; Asp_protease.
DR Pfam; PF00077; rvp; 1.
DR PROSITE; PS00141; ASP_PROTEASE; 1.
DR PROSITE; PS0175; ASP_PROT_RETROV; 1.
KW Hydrolase; Transferase; RNA-directed DNA polymerase;
KW Aspartyl protease; Endonuclease; Polyprotein.
FT ACT_SITE 30
FT NON_TER 128 128 BY SIMILARITY.
SQ SEQUENCE 128 AA; 14116 MW; AC1690EF7378AAC3 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 128;
Best Local Similarity 75.0%; Pred. No. 39;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LPLQLILL 9
Db 118 LPLQLVLT 125

RESULT 31
CD8A_BOVIN STANDARD; PRT; 242 AA.
ID CD8A_BOVIN
AC P31783;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein CD8 alpha chain precursor.
GN CD8A.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Thymus;
RX MEDLINE=92332098; PubMed=1628904;
RA Lalor P., Bucci C., Fornaro M., Rattazzi M.C., Nakauchi H.,
RA Herzenberg L.A., Alberti S.;
RT "Molecular cloning, reconstruction and expression of the gene
RT encoding the alpha-chain of the bovine CD8 -- definition of three
RT peptide regions conserved across species.";
RL Immunology 76:95-102(1992).
CC -!- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT
CC WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN
CC THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO
CC CLASS I MHC MOLECULES ALPHA-3 DOMAINS.
CC -!- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN
CC LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORMS HOMODIMERS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X59416; CAA42051.1; -.
CC PIR; S25663; S25663.
CC HSSP; P01732; 1CD8.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_v.
DR Pfam; PF00047; ig; 1.
DR SMART; SM00406; IGV; 1.
DR Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell; MHC I;
KW Signal.
FT SIGNAL 1 25 POTENTIAL.
FT CHAIN 26 242 T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA
```

```

FT DOMAIN 26 189 CHAIN.
FT TRANSHEM 190 214 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 215 242 POTENTIAL.
FT DOMAIN 26 140 CYTOPLASMIC (POTENTIAL).
FT DISULFID 47 120 IG-LIKE V-TYPE DOMAIN.
SQ SEQUENCE 242 AA; 26417 MW; 91481320EF05195E CRC64;

Query Match 75.0%; Score 30; DB 1; Length 242;
Best Local Similarity 66.7%; Pred. No. 75;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
Db 9 ILPLALLL 17

RESULT 32
YCBC_ECOLI STANDARD; PRT; 259 AA.
AC YCBC_ECOLI
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ycbC.
GN YCBC OR B0920 OR Z1267 OR ECS1003.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamouis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

```

```

RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [5]
RP SEQUENCE OF 1-170 FROM N.A.
RC STRAIN=K12 / W3110;
RX MEDLINE=94232180; PubMed=7513784;
RA Feng J., Yamanaka K., Niki H., Ogura T., Hiraga S.;
RT "New killing system controlled by two genes located immediately
RT upstream of the mukB gene in Escherichia coli.";
RL Mol. Gen. Genet. 243:136-147(1994).
RN [6]
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF000194; AAC74006.1; -
DR EMBL: D90730; BAA35666.1; -
DR EMBL: AE005281; AAG55405.1; -
DR EMBL: AP002553; BAB34426.1; -
DR EMBL: D26440; -; NOT_ANNOTATED_CDS.
DR EcoGene: EG12166; ycbC.
DR InterPro: IPR003848; DUF218.
DR Pfam: PF02698; DUF218; 1.
DR Hypothetical protein; Complete proteome.
FT CONFLICT 89 89 Y -> N (IN REF. 5).
FT CONFLICT 150 171 GVPREQIITLDPKDTDEEAAA -> ACRASKLSPWICQKI
SQ SEQUENCE 259 AA; 28666 MW; 7EAE14C696DAA0C9 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 259;
Best Local Similarity 66.7%; Pred. No. 80;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
Db 13 LLPLPLML 21

RESULT 33
VG74_KSHV STANDARD; PRT; 342 AA.
AC Q98146; O12573; P88966;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE G-protein coupled receptor homolog 7A.
CN 74.
OS Kaposi's sarcoma-associated herpesvirus (KSHV) (Human herpesvirus 8).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=37296;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97048116; PubMed=8892957;
RA Cesarman E., Nador R.G., Bai F., Bohenzky R.A., Russo J.J.,
RA Moore P.S., Chang Y., Knowles D.M.;
RT "Kaposi's sarcoma-associated herpesvirus contains G protein-coupled
RT receptor and cyclin D homologs which are expressed in Kaposi's
RT sarcoma and malignant lymphoma.";
RL J. Virol. 70:8218-8223(1996).
RN [2]
RP REVISIONS.
RA Cesarman E., Nador R.G., Bai F., Chang J., Moore P.S., Knowles D.M.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP COMPLETE GENOME.
RX MEDLINE=97121480; PubMed=8962146;

```

RA Russo J.J., Bohenzky R.A., Chien M.C., Chen J., Yan M., Maddalena D.,
RA Parry J.P., Peruzzi D., Edelman I.S., Chang Y., Moore P.S.:
RA "Nucleotide sequence of the Kaposi sarcoma-associated herpesvirus
RA (HHV8).";
RA Proc. Natl. Acad. Sci. U.S.A. 93:14862-14867(1996).
RL [4]
RN COMPLETE GENOME.
RP MEDLINE=97296220; PubMed=9151804;
RX Neipel F., Albrecht J.-C., Fleckenstein B.:
RA "Cell-homologous genes in the Kaposi's sarcoma-associated
RA rhadinovirus human herpesvirus 8: determinants of its
RT pathogenicity?";
RL J. Virol. 71:4187-4192(1997).
RN [5]
RP SEQUENCE FROM N.A.
RA Guo H.G., Browning P., Nicholas J., Hayward G.H., Tschachler E.,
RA Jiang Y.W., Sadowska M., Raffeld M., Colombini S., Gallo R.C.,
RA Reitz M.S.:
RA Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RL -1- FUNCTION: PROBABLE RECEPTOR FOR A CHEMOKINE. COULD CONTRIBUTE TO
CC THE DEVELOPMENT OF CANCER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U40394; RAD04749.1; -;
DR EMBL; U75698; AAC57160.1; -;
DR EMBL; U93872; AAB62618.1; -;
DR EMBL; U82242; AAB51506.1; -;
DR GCRDB; GCR_1768; -;
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PROSITE; PS00237; G_PROTEIN_RECEP_FL_1; FALSE_NEG.
DR PROSITE; PS50262; G_PROTEIN_RECEP_FL_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 51 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 52 72 1 (POTENTIAL).
FT DOMAIN 73 92 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 93 113 2 (POTENTIAL).
FT DOMAIN 114 121 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 122 142 3 (POTENTIAL).
FT DOMAIN 143 159 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 160 180 4 (POTENTIAL).
FT DOMAIN 181 217 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 218 238 5 (POTENTIAL).
FT DOMAIN 239 251 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 252 272 6 (POTENTIAL).
FT DOMAIN 273 293 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 294 314 7 (POTENTIAL).
FT DOMAIN 315 342 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 18 18 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 22 22 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 31 31 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 44 44 V -> G (IN REF. 2).
FT CONFLICT 167 167 A -> P (IN REF. 2).
FT CONFLICT 170 170 L -> S (IN REF. 5).
FT CONFLICT 322 322 R -> K (IN REF. 5).
SQ SEQUENCE 342 AA; 38668 MW; 41E4B33CA2D9F069 CRC64;
Query Match 75.08; Score 30; DB 1; Length 342;
Best Local Similarity 66.78; Pred No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Oy 1 LLLPQLLL 9
| | | | : | : |
Db 221 LLLPALLL 229

RESULT 34
GPR_HUMAN STANDARD: PRT; 370 AA.
ID GPR_HUMAN
AC P49683; O75194;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Prolactin-releasing peptide receptor (PrRP receptor) (G protein-
DE coupled receptor GPR10) (hGRJ).
GN GPR10 OR GR3.
OS Homo sapiens (Human).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96115583; PubMed=8666380;
RA Marchese A., Heiber M., Nguyen T., Heng H.H.O., Saldivia V.R.,
RA Cheng R., Murphy P.M., Tsui L.-C., Shi X., Gregor P., George S.R.,
RA O'Dowd B.F., Docherty J.M.:
RT "Cloning and chromosomal mapping of three novel genes, GPR9, GPR10,
RT Y, and somatostatin receptors related to interleukin 8, neuropeptide
RL Genomics 29:335-344(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98268781; PubMed=9607765;
RA Hinuma S., Habata Y., Fujii R., Kawamata Y., Hosoya M., Fukusumi S.,
RA Kitada C., Masuo Y., Asano T., Matsumoto H., Sekiguchi M.,
RA Kurokawa T., Nishimura O., Onda H., Fujino M.:
RT "A prolactin-releasing peptide in the brain";
RL Nature 393:272-276(1998).
RN [3]
RP TISSUE SPECIFICITY.
RX MEDLINE=99426652; PubMed=10498338;
RA Fujii R., Fukusumi S., Hosoya M., Kawamata Y., Habata Y., Hinuma S.,
RA Sekiguchi M., Kitada C., Kurokawa T., Nishimura O., Onda H.,
RA Sumino Y., Fujino M.:
RT "Tissue distribution of prolactin-releasing peptide (PrRP) and its
RT receptor";
RL Regul. Pept. 83:1-10(1999).
CC -1- FUNCTION: Receptor for prolactin-releasing peptide (PrRP).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- TISSUE SPECIFICITY: Only detected in the pituitary gland.
CC -1- SIMILARITY: BELONGS TO FAMILY 1 OF G-PROTEIN COUPLED RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32672; AAC50504.1; -;
DR EMBL; AB015745; BAA31159.1; -;
DR GCRDB; GCR_1973; -;
DR MIM; 600895; -;
DR InterPro: IPR000276; GPCR_Rhodpsn.
DR Pfam: PF00001; 7tm_1; 1.
DR PRINTS; PS00237; GPCRHHODPSN.
DR PROSITE; PS00237; G_PROTEIN_RECEP_FL_1; 1.
DR PROSITE; PS50262; G_PROTEIN_RECEP_FL_2; 1.
KW G-protein coupled receptor; Transmembrane; Glycoprotein.
FT DOMAIN 1 62 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 63 83 1 (POTENTIAL).
FT DOMAIN 84 101 CYTOPLASMIC (POTENTIAL).
FT TRANSSEM 102 122 2 (POTENTIAL).
FT DOMAIN 123 126 EXTRACELLULAR (POTENTIAL).
FT TRANSSEM 127 147 3 (POTENTIAL).
FT DOMAIN 148 175 CYTOPLASMIC (POTENTIAL).

```

FT TRANSMEM 176 196 4 (POTENTIAL).
FT DOMAIN 197 225 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 226 246 5 (POTENTIAL).
FT DOMAIN 247 276 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 277 297 6 (POTENTIAL).
FT DOMAIN 298 317 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 318 338 7 (POTENTIAL).
FT DOMAIN 339 369 CYTOPLASMIC (POTENTIAL).
FT DISULFID 134 211 BY SIMILARITY.
FT CARBOHYD 27 27 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 36 36 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 89 89 R -> P (IN REF. 1).
FT CONFLICT 170 175 RISRL -> ASRCA (IN REF. 1).
FT CONFLICT 194 194 A -> P (IN REF. 1).
SQ SEQUENCE 370 AA; 41106 MW; 08E9E57FBEGFC809 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 370;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
DB 235 LLPLVILL 243
||||:|

RESULT 35
BM3B_HUMAN
ID BM3B_HUMAN STANDARD; PRT; 478 AA.
AC P55107; O9UCX6;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Bone morphogenetic protein 3b precursor (BMP-3b)
DE (Growth/differentiation factor 10) (GDF-10) (Bone inducing protein)
DE (BIP).
DE GDF10 OR BMP3B.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Femur;
RX MEDLINE=96264636; PubMed=8670277;
RA Hino J., Takao M., Takashita N., Konno Y., Nishizawa T.,
RA Matsuo H., Kangawa K.;
RT "cDNA cloning and genomic structure of human bone morphogenetic
RT protein-3B (BMP-3b)".
RL Biochem. Biophys. Res. Commun. 223:304-310(1996).
RN [2]
RP SEQUENCE OF 360-478 FROM N.A.
RC TISSUE=Uterus;
RX MEDLINE=96115614; PubMed=8679252;
RA Cunningham N.S., Jenkins N.A., Gilbert D.J., Copeland N.G.,
RA Reddi A.H., Lee S.J.;
RT "Growth/differentiation factor-10: a new member of the transforming
RT growth factor-beta superfamily related to bone morphogenetic protein-
RT 3.".
RL Growth Factors 12:99-109(1995).
CC -1- SUBUNIT: HOMODIMER OR HETERODIMER (POTENTIAL).
CC -1- TISSUE SPECIFICITY: FEMUR, BRAIN, LUNG, SKELETAL MUSCLE, PANCREAS
CC AND TESTIS.
CC -1- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
EMBL: D49493; BAA08453.1; -.

```

```

DR EMBL: D49492; BAA08452.1; -.
DR HSP: P12643; 3BMP.
DR MIM: 601361; -.
DR InterPro: IPR001839; TGF-beta.
DR Pfam: PF00019; TGF-beta; 1.
DR ProDom: PD000357; TGF-beta; 1.
DR SMART: SM00204; TGF-beta; 1.
DR PROSITE: PS00250; TGF-BETA_1; 1.
KW Signal; Growth factor; Cytokine; Glycoprotein.
FT SIGNAL 1 33 POTENTIAL.
FT PROPEP 34 368 POTENTIAL.
FT CHAIN 369 478 BONE MORPHOGENETIC PROTEIN 3B.
FT DISULFID 376 443 BY SIMILARITY.
FT DISULFID 405 475 BY SIMILARITY.
FT DISULFID 409 477 BY SIMILARITY.
FT DISULFID 442 442 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 118 118 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 478 AA; 53121 MW; 80AE0FA4C50B23A9 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 478;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
DB 20 LLPLFLLL 28
||||:|

RESULT 36
Y389_CLOPE
ID Y389_CLOPE STANDARD; PRT; 481 AA.
AC P30818;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Possible amino acid transporter CPE0389.
GN CPE0389.
OS Clostridium perfringens.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1502;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=13 / Type A;
RX PubMed=11792842;
RA Shimizu T., Ohtani K., Hirakawa H., Ohshima K., Yamashita A.,
RA Shiba T., Ogasawara N., Hattori M., Kuhara S., Hayashi H.;
RT "Complete genome sequence of Clostridium perfringens, an anaerobic
RT flesh-eater".
RL Proc. Natl. Acad. Sci. U.S.A. 99:996-1001(2002).
RN [2]
RP SEQUENCE OF 1-35 FROM N.A.
RC STRAIN=ATCC 13124;
RX MEDLINE=90212559; PubMed=2108713;
RA van Poelje P.D., Snell E.E.;
RT "Cloning, sequencing, expression, and site-directed mutagenesis of
RT the gene from Clostridium perfringens encoding pyruvoyl-dependent
RT histidine decarboxylase".
RL Biochemistry 29:132-139(1990).
RN [3]
RP SIMILARITY TO OTHER MEMBERS OF THE FAMILY.
RX MEDLINE=93184721; PubMed=6382989;
RA Reizer J., Finley K., Kakada D., McLeod C.L., Reizer A.,
RA Saier M.H. Jr.;
RT "Mammalian integral membrane receptors are homologous to facilitators
RT and antiporters of yeast, fungi, and eubacteria.".
RL Protein Sci. 2:20-30(1993).
CC -1- FUNCTION: COULD BE AN AMINO ACID TRANSPORTER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE ARCD/CAOB/GERAB/LYSI FAMILY OF

```

PRMSEASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AP003186; BAB80095.1; -
DR EMBL; J02880; AAA83527.1; -
DR PR; B33770; B33770.
KW Transport; Amino-acid transport; Transmembrane; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 36 56 POTENTIAL.
FT TRANSMEM 87 107 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 208 228 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 289 309 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 364 384 POTENTIAL.
FT TRANSMEM 401 421 POTENTIAL.
FT TRANSMEM 422 442 POTENTIAL.
FT TRANSMEM 461 481 POTENTIAL.
SQ SEQUENCE 481 AA; 52759 MW; 8C2910FDE2A91F CRC64;

Query Match 75.0%; Score 30; DB 1; Length 481;

Best Local Similarity 77.8%; Pred. No. 1.5e+02;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9

Db 162 LLPLAIFLL 170

RESULT 37

TTDT_ECOLI

ID P39414; Q46870; STANDARD; PRT; 487 AA.

AC 01-FEB-1995 (Rel. 31, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Putative tartrate carrier (Tartrate transporter) (Tartrate/succinate

DE antiporter).

GN YGJE OR B3063.

OS Escherichia coli.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562;

FN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RL Science 277:1453-1474(1997).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE=87248073; PubMed=3297921;

RA Nesin M., Lupski J.R., Svec P., Godson G.N.;

RT "Possible new genes as revealed by molecular analysis of a 5-kb

RT Escherichia coli chromosomal region 5' to the rpsU-dnaG-rpoD

RT macromolecular-synthesis operon.";

RL Gene 51:149-161(1987).

RN [3]

RP IDENTIFICATION.

RX MEDLINE=95075659; PubMed=7984428;

RA Borodovsky M., Rudd K.E., Koonin E.V.;

RT "Intrinsic and extrinsic approaches for detecting genes in a

RT bacterial genome.";

RL Nucleic Acids Res. 22:4756-4767(1994).

RN [4]

RP POSSIBLE FUNCTION.

RX POS K.M., Dimroth P., Bott M.;

RT "The Escherichia coli citrate carrier Ctt: a member of a novel

RT eubacterial transporter family related to the 2-oxoglutarate/malate

RT translocator from spinach chloroplasts.";

RL J. Bacteriol. 180:4160-4165(1998).

CC -1- FUNCTION: RESPONSIBLE FOR THE UPTAKE OF TARTRATE IN EXCHANGE TO

CC THE EFFLUX OF SUCCINATE (POTENTIAL).

CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane

CC (Probable).

CC -1- SIMILARITY: BELONGS TO THE NADC/P/PHO87 FAMILY OF TRANSPORTERS.

CC SODITI SUBFAMILY.

CC -1- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS

CC IN POSITIONS 35, 51, 132, 245, 268 AND 443.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; U28379; AAA89143.1; -

DR EMBL; AE00388; AAC76099.1; -

DR EMBL; M16194; -; NOT_ANNOTATED_CDS.

DR EcoGene; EG12393; YgJE.

DR InterPro; IPR001898; Na_sulph_syp.

DR Pfam; PF00939; Na_sulph_syp; 1.

KW Hypothetical protein; Transmembrane; Inner membrane; Transport;

KW Complete proteome.

FT TRANSMEM 10 30 POTENTIAL.

FT TRANSMEM 33 53 POTENTIAL.

FT TRANSMEM 54 74 POTENTIAL.

FT TRANSMEM 93 113 POTENTIAL.

FT TRANSMEM 137 157 POTENTIAL.

FT TRANSMEM 189 209 POTENTIAL.

FT TRANSMEM 236 256 POTENTIAL.

FT TRANSMEM 292 312 POTENTIAL.

FT TRANSMEM 313 333 POTENTIAL.

FT TRANSMEM 340 360 POTENTIAL.

FT TRANSMEM 370 390 POTENTIAL.

FT TRANSMEM 393 413 POTENTIAL.

FT TRANSMEM 418 438 POTENTIAL.

FT TRANSMEM 465 485 POTENTIAL.

FT CONFLICT 404 404 L -> P (IN REF. 2).

FT CONFLICT 457 457 A -> T (IN REF. 2).

SQ SEQUENCE 487 AA; 52906 MW; 4D6E8439BEE281D2 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 487;

Best Local Similarity 66.7%; Pred. No. 1.5e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9

Db 239 MPLSLILV 247

RESULT 38

NU4M_CHOCR

ID NU4M_CHOCR STANDARD; PRT; 492 AA.

AC P48915;

DT 01-FEB-1996 (Rel. 33, Created)

DT 01-FEB-1996 (Rel. 33, Last sequence update)

DT 01-FEB-1996 (Rel. 33, Last annotation update)

DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3).

GN ND4 OR NAD4.

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----
DR EMBL; AB029017; BAA83046.1; -
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 14 34 POTENTIAL.
FT TRANSMEM 75 95 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 255 275 POTENTIAL.
FT TRANSMEM 298 318 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT TRANSMEM 402 422 POTENTIAL.
FT TRANSMEM 437 457 POTENTIAL.
FT TRANSMEM 473 493 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
SQ SEQUENCE 538 AA; 59267 MW; EB7D1BABD45362AD CRC64;

Query Match 75.0%; Score 30; DB 1; Length 538;

Best Local Similarity 66.7%; Pred. No. 1.7e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9

||||:|

Db 527 LLPLYLL 535

RESULT 41

SOHC_BRAJA

ID SOHC_BRAJA STANDARD; PRT; 658 AA.

AC P34924;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Squalene-hopene cyclase (EC 5.4.99.-)

GN SHC.

OS Bradyrhizobium japonicum.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Bradyrhizobium group; Bradyrhizobium.

OX NCBI_TaxID=375;

RN [1]

RP SEQUENCE FROM N.A.

Perzi M., Poralla K.;

RL Submitted (OCT-1995) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: CATALYZES THE CYCLIZATION OF SQUALENE INTO HOPENE.

CC -1- PATHWAY: KEY ENZYME IN HOPANOID (TRITERPENOID) METABOLISM.

CC -1- SUBCELLULAR LOCATION: Membrane-associated (By similarity).

CC -1- SIMILARITY: BELONGS TO THE TERPENE CYCLASE/MUTASE FAMILY.

CC -1- SIMILARITY: CONTAINS 3 PFTB REPEATS.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; X86552; CAA60250.1; -

DR HSSP; P33247; 25QC.

DR InterPro; IPR001330; Prenyltrans.

DR InterPro; IPR002365; Terpene_synth.

DR Pfam; PF00432; prenyltrans; 4.

DR PROSITE; PS01074; TERPENE_SYNTHASES; 1.

KW Isomerase; Membrane; Repeat. PFTB 1.

FT REPEAT 71 112

FT REPEAT 417 458 PFTB 2.

FT REPEAT 534 584 PFTB 3.

SQ SEQUENCE 658 AA; 73412 MW; 57C7E58F42783A8B CRC64;

Query Match 75.0%; Score 30; DB 1; Length 658;

Best Local Similarity 55.6%; Pred. No. 2.1e+02;

Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9

||||:|

Db 157 VLPIEIMLL 165

RESULT 42

GARN_DROME

ID GARN_DROME STANDARD; PRT; 810 AA.

AC P54362;

DT 01-OCT-1996 (Rel. 34, Created)

DT 15-JUL-1999 (Rel. 38, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE Garnet protein.

GN G.

OS Drosophila melanogaster (Fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Prorygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OREGON-R;

RA Lloyd V.K.;

RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO THE ADAPTOR COMPLEXES LARGE SUBUNIT

CC FAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL; U31351; AAB97618.1; -

DR FlyBase; FBgn001087; g.

DR InterPro; IPR002553; Adaptin_N.

DR Pfam; PF01602; Adaptin_N; 1.

SQ SEQUENCE 810 AA; 90199 MW; C1941DB73D1EA262 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 810;

Best Local Similarity 85.7%; Pred. No. 2.6e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 PLQILL 9

||||:|

Db 74 PLQILL 80

RESULT 43

MPRI_BOVIN

ID MPRI_BOVIN STANDARD; PRT; 2499 AA.

AC P08169;

DT 01-AUG-1988 (Rel. 08, Created)

DT 01-OCT-1989 (Rel. 12, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Cation-independent mannose-6-phosphate receptor precursor (CI MAN-6-P

DE receptor) (CI-MPR) (Insulin-like growth factor II receptor) (300 KDA
DE mannose 6-phosphate receptor) (MPR 300) (MPR300).

GN IGF2R OR M6P.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;

OC Bovidae; Bovinae; Bos.

OX NCBI_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=88115411; PubMed=2963004;


```

DR Pfam; PF00616; RasGAP; 1.
DR SMART; SM00323; RasGAP; 1.
DR SMART; SM00516; SEC14; 1.
DR PROSITE; PS00509; RAS_GTPASE_ACTIV_1; 1.
DR PROSITE; PS00018; RAS_GTPASE_ACTIV_2; 1.
KW GTPase activation.
FT DOMAIN 1237 1432 RAS-GAP.
FT DOMAIN 1354 1357 POLY-SER.
SQ SEQUENCE 2820 AA; 317080 MW; FC108487E86DA89F CRC64;

Query Match 75.0%; Score 30; DB 1; Length 2820;
Best Local Similarity 85.7%; Pred. No. 9.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 PLOIILL 9
Db 268 PLOIILL 274
||||:|

RESULT 45
NF1_HUMAN STANDARD; PRT: 2839 AA.
ID NF1_HUMAN STANDARD; Q14284; Q14930; Q9UMK3;
AC P21359; O00662; Q14284; Q14930; Q9UMK3;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Neurofibromin (Neurofibromatosis-related protein NF-1).
GN NF1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE-93090270; PubMed=1457041;
RA Bernards A., Haase V.H., Murthy A.E., Menon A., Hannigan G.E.,
RA Gusella J.F.;
RT "Complete human NF1 cDNA sequence: two alternatively spliced mRNAs
RT and absence of expression in a neuroblastoma line.";
RL DNA Cell Biol. 11:727-734(1992).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE-90319792; PubMed=2134734;
RA Wallace M.R., Marchuk D.A., Andersen L.B., Letcher R., Odeh H.M.,
RA Saullino A.M., Fountain J.W., Brereton A., Nicholson J., Mitchell A.L.,
RA Brownstein B.H., Collins F.S.;
RT "Type 1 neurofibromatosis gene: identification of a large transcript
RT disrupted in three NF1 patients.";
RL Science 249:181-186(1990).
RN [3]
RP ERRATUM.
RX MEDLINE-91102559; PubMed=2125369;
RA Wallace M.R., Marchuk D.A., Andersen L.B., Collins F.S.;
RL Science 250:1749-1749(1990).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE-92147138; PubMed=1783401;
RA Marchuk D.A., Saullino A.M., Tavakoli R., Swaroop M., Wallace M.R.,
RA Andersen L.B., Mitchell A.L., Gutmann D.H., Boguski M.S.,
RA Collins F.S.;
RT "cDNA cloning of the type 1 neurofibromatosis gene: complete sequence
RT of the NF1 gene product.";
RL Genomics 11:931-940(1991).
RN [5]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RX MEDLINE-92412152; PubMed=1339276;
RA Suzuki H., Takahashi K., Kubota Y., Shibahara S.;
RT "Molecular cloning of a cDNA coding for neurofibromatosis type 1
RT protein isoform lacking the domain related to ras GTPase-activating
RT protein.";
RL Biochem. Biophys. Res. Commun. 187:984-990(1992).
RN [6]
RP SEQUENCE OF 335-2839 FROM N.A. (ISOFORM 1), AND VARIANT PRO-1953.
RX MEDLINE-90304909; PubMed=2114220;
RA Cawthon R.M., Weiss R., Xu G., Viskochil D., Culver M., Stevens J.,
RA Robertson M., Dunn D., Gesteland R., O'Connell P., White R.;
RT "A major segment of the neurofibromatosis type 1 gene: cDNA sequence,
RT genomic structure, and point mutations.";
RL Cell 62:193-201(1990).
RN [7]
RP SEQUENCE OF 335-2839 FROM N.A. (ISOFORM 1).
RX MEDLINE-90335969; PubMed=2116237;
RA Xu G., O'Connell P., Viskochil D., Cawthon R.M., Robertson M.,
RA Culver M., Dunn D., Stevens J., Gesteland R., White R., Weiss R.;
RT "The neurofibromatosis type 1 gene encodes a protein related to GAP.";
RL Cell 62:599-608(1990).
RN [8]
RP SEQUENCE OF 707-782 FROM N.A.
RX MEDLINE-97156204; PubMed=9002664;
RA Regnier V., Meddeb M., Lecointre G., Richard F., Duverger A.,
RA Nguyen V.C., Dutrillaux B., Bernheim A., Dangelot G.;
RT "Emergence and scattering of multiple neurofibromatosis (NF1)-related
RT sequences during hominoid evolution suggest a process of
RT pericentromeric interchromosomal transposition.";
RL Hum. Mol. Genet. 6:9-16(1997).
RN [9]
RP SEQUENCE OF 751-1611 FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE-95293414; PubMed=7774960;
RA Li Y., O'Connell P., Breidenbach H.H., Cawthon R.M., Stevens J.,
RA Xu G., Neil S., Robertson M., White R., Viskochil D.;
RT "Genomic organization of the neurofibromatosis 1 gene (NF1).";
RL Genomics 25:9-18(1995).
RN [10]
RP SEQUENCE OF 1090-1598 FROM N.A. (ISOFORM 4).
RX MEDLINE-91029515; PubMed=2121370;
RA Martin G.A., Viskochil D., Bollag G., McCabe P.C., Crosier W.J.,
RA Haubruck H., Conroy L., Clark R., O'Connell P., Cawthon R.M.,
RA Innis M., McCormick F.;
RT "The GAP-related domain of the neurofibromatosis type 1 gene product
RT interacts with ras p21.";
RL Cell 63:843-849(1990).
RN [11]
RP SEQUENCE OF 1168-1566 FROM N.A. (ISOFORMS 1 AND 2).
RX MEDLINE-92019823; PubMed=1923522;
RA Nishi T., Lee P.S., Oka K., Levin V.A., Tanase S., Morino Y.,
RA Sava H.;
RT "Differential expression of two types of the neurofibromatosis type 1
RT (NF1) gene transcripts related to neuronal differentiation.";
RL Oncogene 6:1555-1559(1991).
RN [12]
RP SEQUENCE OF 1371-1391 FROM N.A. (ISOFORM 2).
RX MEDLINE-93109335; PubMed=8417346;
RA Andersen L.B., Ballester R., Marchuk D.A., Chang E., Gutmann D.H.,
RA Saullino A.M., Camonis J., Wigler M., Collins F.S.;
RT "A conserved alternative splice in the von Recklinghausen
RT neurofibromatosis (NF1) gene produces two neurofibromin isoforms,
RT both of which have GTPase-activating protein activity.";
RL Mol. Cell. Biol. 13:487-495(1993).
RN [13]
RP SEQUENCE OF 1371-1391 FROM N.A. (ISOFORM 2).
RX MEDLINE-92109786; PubMed=1662505;
RA Suzuki Y., Suzuki H., Kayama T., Yoshimoto T., Shibahara S.;
RT "Brain tumors predominantly express the neurofibromatosis type 1
RT gene transcripts containing the 63 base insert in the region coding
RT for GTPase activating protein-related domain.";
RL Biochem. Biophys. Res. Commun. 181:955-961(1991).
RN [14]
RP FUNCTION.
RX MEDLINE-91029516; PubMed=2121371;
RA Ballester R., Marchuk D.A., Boguski M.S., Saullino A.M., Letcher R.,
RA Wigler M., Collins F.S.;
RT "The NF1 locus encodes a protein functionally related to mammalian
RT GAP and yeast Ira proteins.";
RL Cell 63:851-859(1990).
RN [15]
RP REVIEW ON VARIANTS.

```

RX MEDLINE-95072625; PubMed-7981724;
RA Upadhyaya M., Shaw D.J., Harper P.S.;
RT "Molecular basis of neurofibromatosis type 1 (NF1): mutation analysis
RL and polymorphisms in the NF1 gene.";
RN Hum. Mutat. 4:83-101(1994).
[16]
RP REVIEW ON VARIANTS.
RX MEDLINE-96422425; PubMed-8825042;
RA Shen M.H., Harper P.S., Upadhyaya M.;
RT "Molecular genetics of neurofibromatosis type 1 (NF1).";
RL J. Med. Genet. 33:2-17(1996).
[17]
RP VARIANT GLU-1444.
RX MEDLINE-92233464; PubMed-1568247;
RA Li Y., Bollag G., Clark R., Stevens J., Conroy L., Fults D., Ward K.,
RA Friedman E., Samowitz W., Robertson M., Bradley P., McCormick F.,
RA White R., Cawthon R.M.;
RT "Somatic mutations in the neurofibromatosis 1 gene in human tumors.";
RL Cell 69:275-281(1992).
[18]
RP VARIANTS MET-2164 AND ASN-2192.
RX MEDLINE-93258316; PubMed-1302608;
RA Upadhyaya M., Shen M.H., Cherryson A., Farnham J., Maynard J.,
RA Huson S.M., Harper P.S.;
RT "Analysis of mutations at the neurofibromatosis 1 (NF1) locus.";
RL Hum. Mol. Genet. 1:735-740(1992).
[19]
RP VARIANT HIS-1721--LEU-1733 DUPL.
RX MEDLINE-93304433; PubMed-8317503;
RA Tassabehji M., Strachan T., Sharland M., Colley A., Donnai D.,
RA Harris R., Thakker N.;
RT "Tandem duplication within a neurofibromatosis type 1 (NF1) gene exon
RL in a family with features of Watson syndrome and Noonan syndrome.";
RN Am. J. Hum. Genet. 53:90-95(1993).
[20]
RP VARIANT MET-991 DEL.
RX MEDLINE-94108439; PubMed-7904209;
RA Shen M.H., Harper P.S., Upadhyaya M.;
RT "Neurofibromatosis type 1 (NF1): the search for mutations by PCR-
RL heteroduplex analysis on Hydrolink gels.";
RN Hum. Mol. Genet. 2:1861-1864(1993).
[21]
RP VARIANTS NF1 ASP-1166 AND ARG-1440.
RX MEDLINE-95072575; PubMed-7981679;
RA Purandare S.M., Lanyon W.G., Connor J.M.;
RT "Characterisation of inherited and sporadic mutations in
RL neurofibromatosis type-1.";
RN Hum. Mol. Genet. 3:1109-1115(1994).
[22]
RP VARIANT NF1 2387-ASN-PHE-2388 DEL.
RX MEDLINE-94362704; PubMed-8081387;
RA Abernathy C.R., Colman S.D., Kousseff B.G., Wallace M.R.;
RT "Two NF1 mutations: frameshift in the GAP-related domain, and loss of
RL two codons toward the 3' end of the gene.";
RN Hum. Mutat. 3:347-352(1994).
[23]
RP VARIANT NF1 ALA-2631.
RX MEDLINE-96091873; PubMed-8544190;
RA Upadhyaya M., Maynard J., Osborn M.J., Huson S.M., Ponder M.,
RA Ponder B.A.J., Harper P.S.;
RT "Characterisation of germline mutations in the neurofibromatosis type
RL 1 (NF1) gene.";
RN J. Med. Genet. 32:706-710(1995).
[24]
RP VARIANT NF1 ARG-629.
RX MEDLINE-96431167; PubMed-8834249;
RA Gasparini P., D'Agruma L., de Cillis G.P., Balestrazzi P.,
RA Mingarelli R., Zelante L.;
RT "Scanning the first part of the neurofibromatosis type 1 gene by RNA-
RL SSCP: identification of three novel mutations and of two new
RN polymorphisms";
RN Hum. Genet. 97:492-495(1996).
[25]

RP VARIANT NF1 ARG-1035.
RX MEDLINE-96400960; PubMed-8807336;
Query Match 75.0%; Score 30; DB 1; Length 2839;
Best Local Similarity 85.7%; Pred. No. 9.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 PLQILL 9
Db 268 PLQILL 274
Search completed: November 6, 2002, 12:09:13
Job time : 9.33333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 11 Seconds
(without alignments)
78.619 Million cell updates/sec

Title: US-09-905-083-35
Perfect score: 40
Sequence: 1 LLPLQILL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	40	100.0	253	A53968	serine proteinase
2	36	90.0	845	D97163	cation transport p
3	35	87.5	493	A71875	hypothetical prote
4	35	87.5	496	B64638	conserved hypothet
5	34	85.0	266	JC7300	tax-responsive ele
6	34	85.0	267	JC4857	hepatocarcinogenes
7	34	85.0	369	S33603	surfactant protein
8	34	85.0	371	JN0450	conglutinin precu
9	34	85.0	371	I45878	conglutinin - bovi
10	34	85.0	429	T17215	hypothetical prote
11	34	85.0	539	H82994	probable permease
12	34	85.0	1092	T12520	hypothetical prote
13	34	85.0	1236	T50904	Mg protoporpyrin
14	33	82.5	146	H75201	hypothetical prote
15	33	82.5	277	H84314	cytochrome aa3 con
16	33	82.5	282	E70890	hypothetical prote
17	33	82.5	294	A25970	transcription acti
18	33	82.5	294	E82257	cholera toxin tran
19	33	82.5	653	A46362	amyloid precursor-
20	32	80.0	245	E75440	hypothetical prote
21	32	80.0	253	AF1535	transport protein
22	32	80.0	256	JQ0708	lipopolysaccharide
23	32	80.0	394	B43750	chloramphenicol re
24	32	80.0	394	B85535	probable transport
25	32	80.0	394	F90684	probable transport
26	32	80.0	412	A57468	P-selectin glycop
27	32	80.0	491	JC6197	stromelysin 3 (EC
28	32	80.0	492	A44399	stromelysin 3 (EC
29	32	80.0	543	AI2088	Na+/H+-exchanging

30	32	80.0	1041	2	T29010	hypothetical prote
31	32	80.0	1306	1	A31759	peptidyl-di-peptida
32	32	80.0	1485	2	S23756	CFTR protein - Afr
33	31	77.5	157	2	A83214	probable transcrip
34	31	77.5	189	2	T43766	hypothetical prote
35	31	77.5	196	2	G65039	hypothetical prote
36	31	77.5	209	2	A54984	ELF-1 protein prec
37	31	77.5	213	2	JE0322	ephraim-A2 - human
38	31	77.5	303	2	C84914	hypothetical prote
39	31	77.5	306	2	H86811	sugar ABC transport
40	31	77.5	327	1	A41720	acid phosphatase (
41	31	77.5	328	1	A41927	insulin-like growt
42	31	77.5	370	2	AB3334	daunorubicin resis
43	31	77.5	372	2	A98157	probable permease
44	31	77.5	372	2	AH3130	ABC transporter, m
45	31	77.5	388	2	T26360	hypothetical prote

ALIGNMENTS

RESULT 1

A53968
serine proteinase SCCE precursor - human
N;Alternate names: stratum corneum chymotryptic enzyme
C;Species: Homo sapiens (man)
C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
C;Accession: A53968
R;Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, J. Biol. Chem. 269, 19420-19426, 1994
A;Title: Cloning, expression, and characterization of stratum corneum chymotryptic ei
A;Reference number: A53968; MUID:94308225
A;Accession: A53968
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-253 <HAN>
A;Cross-references: GB:I33404; NID:g521214; PIDN:AAC37551.1; PID:g532504
C;Genetics:
A;Gene: GDB:PRSS6; SCCE
A;Cross-references: GDB:377730
A;Map position: 7q35-7q35
C;Superfamily: trypsin; trypsin homology
F;30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.9;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
 | | | | | | | |
DB 6 LLPLQILL 14

RESULT 2

D97163
cation transport p-type ATPase CAC2137 [imported] - Clostridium acetobutylicum
C;Species: Clostridium acetobutylicum
C;Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C;Accession: D97163
R;Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.;
.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
A;Reference number: A96900; MUID:21359325; PMID:21359325
A;Accession: D97163
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-845 <KUR>
A;Cross-references: GB:AE001437; PIDN:AAK80095.1; PID:gl15025128; GSPDB:GN00168
A;Experimental source: Clostridium acetobutylicum ATCC824
C;Genetics:
A;Gene: CAC2137

Query Match 90.0%; Score 36; DB 2; Length 845;
Best Local Similarity 77.8%; Pred. No. 40;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
|||||
Db 670 LLPLQIILL 678

RESULT 3

A:1875

hypothetical protein jhp0880 - Helicobacter pylori (strain J99)

C:Species: Helicobacter pylori

A:Variety: strain J99

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999

C:Accession: A71875

R:Alm, R.A.; Ling, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doig, P.C.; Smith, D.R.;
Ives, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Vovis, G.F.;
Nature 397, 176-180, 1999

A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path

A:Reference number: A71800; MUID:99120557

A:Accession: A71875

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-493 <ARN>

A:Cross-references: GB:AE001518; GB:AE001439; NID:g4155454; PIDN:AAD06464.1; PID:g415546

A:Experimental source: strain J99

C:Genetics:

A:Gene: jhp0880

Query Match 87.5%; Score 35; DB 2; Length 493;
Best Local Similarity 88.9%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
|||||
Db 254 LLPLSILL 262

RESULT 4

B64638

conserved hypothetical integral membrane protein HP0946 - Helicobacter pylori (strain 26

C:Species: Helicobacter pylori

C:Date: 09-Aug-1997 #sequence_revision 09-Aug-1997 #text_change 08-Oct-1999

C:Accession: B64638

R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khaliak, H.G.; Glodek, A.; McKenne

son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997

A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.

A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.

A:Reference number: A64520; MUID:97394467

A:Accession: B64638

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-496 <TOM>

A:Cross-references: GB:AE000604; GB:AE000511; NID:g2314086; PIDN:AAD07993.1; PID:g231408

C:Genetics:

A:Start codon: GTG

Query Match 87.5%; Score 35; DB 2; Length 496;
Best Local Similarity 88.9%; Pred. No. 36;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
|||||
Db 254 LLPLSILL 262

RESULT 5

JC7300

tax-responsive element-binding protein 5 - mouse

C:Species: Mus musculus (house mouse)

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C:Accession: JC7300

R:Masaki, T.; Noguchi, H.; Kobayashi, M.; Yoshida, M.; Takamatsu, K.

DNA Res. 7, 187-193, 2000

A:Title: Isolation and characterization of the gene encoding mouse tax-responsive ele

A:Reference number: JC7300

A:Accession: JC7300

A:Molecule type: mRNA

A:Residues: 1-266 <MAS>

A:Cross-references: DDBJ:AB036745

C:Genetics:

A:Gene: treb5

C:Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C:Keywords: leucine zipper; transcription factor

Query Match 85.0%; Score 34; DB 2; Length 266;

Best Local Similarity 88.9%; Pred. No. 31;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
|||||
Db 187 LLPLQIILL 195

RESULT 6

JC4857

hepatocarcinogenesis-related transcription factor - rat

C:Species: Rattus norvegicus (Norway rat)

C:Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 08-Sep-1997

C:Accession: JC4857

R:Kishimoto, T.; Kokura, K.; Kumagai, Y.; Makino, Y.; Tamura, T.

Biochem. Biophys. Res. Commun. 224, 746-751, 1996

A:Title: HFE: A b-zip transcription factor that is closely related to the human XBP/T

A:Reference number: JC4857

A:Accession: JC4857

A:Molecule type: mRNA

A:Residues: 1-267 <KTS>

C:Comment: This is a basic-leucine zipper type transcription factor involved in hepat

C:Genetics:

A:Gene: htf

C:Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C:Keywords: leucine zipper; transcription factor

F:58-98/Domain: fos/jun DNA-binding domain homology <FUD>

F:89-126/Region: leucine zipper motif

Query Match 85.0%; Score 34; DB 2; Length 267;

Best Local Similarity 88.9%; Pred. No. 31;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
|||||
Db 188 LLPLQIILL 196

RESULT 7

S33603

surfactant protein D - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999

C:Accession: S33603

R:Llm, B.L.; Lu, J.; Reid, K.B.M.

Immunology 78, 159-165, 1993

A:Title: Structural similarity between bovine conglutinin and bovine lung surfactant

A:Reference number: S33603; MUID:93170856

A:Accession: S33603

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-369 <LIM>

C:Superfamily: pulmonary surfactant protein D; C-type lectin homology

F:248-367/Domain: C-type lectin homology <LCH>

Query Match 85.0%; Score 34; DB 2; Length 369;

Best Local Similarity 77.8%; Pred. No. 42;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQIILL 9
||||:|

Db 3 LLPLSVLL 11

RESULT 8

JN0450

conglutinin precursor - bovine

N;Alternate names: C3b-binding protein

C;Contains: conglutinin-N

C;Species: Bos primigenius taurus (cattle)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C;Accession: JN0450; JC3235; A23740; S36879; S35044; I46010; A29416; S34054

R;Suzuki, Y.; Yin, Y.P.; Makino, M.; Kurimura, T.; Wakamiya, N.

Biochem. Biophys. Res. Commun. 191, 335-342, 1993

A;Title: Cloning and sequencing of a cDNA coding for bovine conglutinin.

A;Reference number: JN0450; MUID:93213261

A;Accession: JN0450

A;Molecule type: mRNA

A;Residues: 1-371 <SUZ>

A;Cross-references: DDBJ:D14085; NID:9285643; PIDN:BA03170.1; PID:g285644

A;Experimental source: liver

R;Kawasaki, N.; Itoh, N.; Kawasaki, T.

Biochem. Biophys. Res. Commun. 198, 597-604, 1994

A;Title: Gene organization and 5'-flanking region sequence of conglutinin: A C-type mann

A;Reference number: JC2396; MUID:94128104

A;Accession: JC2396

A;Molecule type: mRNA

A;Residues: 1-371 <KA2>

A;Note: The authors translated the codon GAT for residues 250 and 270 as Glu

R;Lu, J.; Laursen, S.B.; Thiel, S.; Jensenius, J.C.; Reid, K.B.M.

Biochem. J. 292, 157-162, 1993

A;Title: The cDNA cloning of conglutinin and identification of liver as a primary site of

A;Reference number: S33235; MUID:93277452

A;Accession: S33235

A;Molecule type: mRNA

A;Residues: 1-172, 'H', 174-217, 'A', 219-271, 'V', 273-371 <LUJ>

A;Cross-references: EMBL:X1774; NID:g395267; PIDN:CA050665.1; PID:g395268

A;Experimental source: liver

R;Lee, Y.M.; Leiby, K.R.; Allar, J.; Paris, K.; Lerch, B.; Okarma, T.B.

J. Biol. Chem. 266, 2715-2723, 1991

A;Title: Primary structure of bovine conglutinin, a member of the C-type animal lectin f

A;Reference number: A23740; MUID:91131556

A;Accession: A23740

A;Molecule type: protein

A;Residues: 21-209, 'S', 211-371 <LEE>

R;Kawasaki, N.; Yokota, Y.; Kawasaki, T.

Arch. Biochem. Biophys. 305, 533-540, 1993

A;Title: Differentiation of conglutination activity and sugar-binding activity of conglu

A;Reference number: S36879; MUID:93384312

A;Accession: S36879

A;Molecule type: protein

A;Residues: 21-54; 75-86, 'X', 88-89, 'X', 91, 'X', 93-94; 208-209, 'X', 211-227 <KAW>

A;Experimental source: serum

R;Lu, J.; Wiedemann, H.; Holmskov, U.; Thiel, S.; Timpl, R.; Reid, K.B.M.

Eur. J. Biochem. 215, 793-799, 1993

A;Title: Structural similarity between lung surfactant protein D and conglutinin. Two di

A;Reference number: S35044; MUID:93358905

A;Accession: S35044

A;Molecule type: protein

A;Residues: 75-86, 'X', 88-89, 'X', 91, 'I' <LUA>

A;Experimental source: lung

R;Young, N.M.; Leon, M.A.

Biochem. Biophys. Res. Commun. 143, 645-651, 1987

A;Title: The carbohydrate specificity of conglutinin and its homology to proteins in the

A;Reference number: A29416; MUID:87184551

A;Contents: annotation

R;Malhotra, R.; Laursen, S.B.; Willis, A.C.; Sim, R.B.

Biochem. J. 293, 15-19, 1993

A;Title: Research Communication. Localization of the receptor-binding site in the collec

A;Reference number: S34054; MUID:93319501

A;Contents: annotation
R;Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry, J. Immunol. 153, 173-180, 1994
A;Title: Bovine conglutinin gene exon structure reveals its evolutionary relationship
A;Reference number: I46010; MUID:94267222
A;Accession: I46010
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-371 <LIO>

A;Cross-references: EMBL:U06860; NID:g507183; PIDN:AAB60624.1; PID:g514256
C;Comment: This protein mediates the agglutination of erythrocytes with antibody and
C;Genetics:
A;Gene: CGN1

A;Introns: 67/1; 106/1; 142/1; 178/1; 217/1; 245/1

C;Superfamily: pulmonary surfactant protein D; C-type lectin homology

C;Keywords: calcium binding; glycoprotein; hydroxylysine; hydroxyproline

F;1-20/Domain: signal sequence #status predicted <SIG>

F;21-371/Product: conglutinin #status predicted <MAT>

F;46-214/Region: collagen-like

F;75-371/Product: conglutinin-N #status predicted <MA2>

F;248-369/Domain: C-type lectin homology <LCH>

F;63,87,99,135,141,159,162,198,210/Binding site: carbohydrate (Lys) (covalent) #stat

F;78,96,108,111,129,132,147,153,171,195/Modified site: 5-hydroxylysine (Lys) #status expe

F;337/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 85.0%; Score 34; DB 1; Length 371;

Best Local Similarity 77.8%; Pred. No. 43;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQIILL 9

Db 3 LLPLSVLL 11

RESULT 9

I45878

conglutinin - bovine

C;Species: Bos primigenius taurus (cattle)

C;Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 20-Aug-1999

C;Accession: I45878

R;Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sastry, J. Immunol. 153, 173-180, 1994

A;Title: Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterizati

A;Reference number: I45878; MUID:94215917

A;Accession: I45878

A;Status: preliminary; translated from GB/EMBL/DBJ

A;Molecule type: mRNA

A;Residues: 1-371 <LIO>

A;Cross-references: GB:LI8871; NID:g495012; PIDN:AAA20126.1; PID:g495013

C;Superfamily: pulmonary surfactant protein D; C-type lectin homology

F;248-369/Domain: C-type lectin homology <LCH>

Query Match 85.0%; Score 34; DB 2; Length 371;

Best Local Similarity 77.8%; Pred. No. 43;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQIILL 9

Db 3 LLPLSVLL 11

RESULT 10

T17215

hypothetical protein DKFp434H2235.1 - human

C;Species: Homo sapiens (man)

C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C;Accession: T17215

R;Poustka, A.; Klein, M.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, September 1999

A;Reference number: 218723

A;Accession: T17215

A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-429 <POU>
A:Cross-references: EMBL:AL117404
A:Experimental source: adult testis; clone DKFZp434H2235
C:Genetics:
A:Note: DKFZp434H2235.1

Query Match 85.0%; Score 34; DB 2; Length 429;
Best Local Similarity 87.5%; Pred. No. 49;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
|||||
Db 381 LPLQLLLL 388

RESULT 11

H82994
probable permease of ABC iron transporter PA5216 [imported] - Pseudomonas aeruginosa (strain PAO1)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: H82994
R:Stover, C.K.; Phan, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Boman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PAO1, an opportunistic pathogen
A:Reference number: A82950; MUID:20437337
A:Accession: H82994
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-539 <STO>
A:Cross-references: GB:AE004934; GB:AE004091; NID:9951515; PIDN:AAG08601.1; GSPDB:GN001
A:Experimental source: strain PAO1
C:Genetics:
A:Gene: PA5216
C:Superfamily: sfuB protein

Query Match 85.0%; Score 34; DB 2; Length 539;
Best Local Similarity 77.8%; Pred. No. 62;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPLQILL 9
|||||
Db 21 LPLSVLL 29

RESULT 12

T12520
hypothetical protein DKFZp434G173.1 - human
C:Species: Homo sapiens (man)
C:Date: 23-Jul-1999 #sequence_revision 23-Jul-1999 #text_change 23-Jul-1999
C:Accession: T12520
R:Wambutt, R.; Heubner, D.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, June 1999
A:Reference number: Z17524
A:Accession: T12520
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1092 <KAW>
A:Cross-references: EMBL:AL080133
A:Experimental source: adult testis; clone DKFZp434G173
C:Genetics:
A:Note: DKFZp434G173.1

Query Match 85.0%; Score 34; DB 2; Length 1092;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
|||||
Db 1044 LPLQLLLL 1051

RESULT 13

T50904
Mg protoporphyrin methyl transferase [imported] - Rubrivivax gelatinosus
C:Species: Rubrivivax gelatinosus
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 02-Sep-2000
C:Accession: T50904
R:Nagashima, K.V.; Igarashi, N.; Harada, J.; Nagashima, S.; Matsuura, K.; Shimada, K.
submitted to the EMBL Data Library, November 1999
A:Description: Determination of Nucleotide Sequences of Rubrivivax gelatinosus Photosynthesis Genes
A:Reference number: Z25270
A:Accession: T50904
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1236 <NAG>
A:Cross-references: EMBL:AB034704; PIDN:BAA94057.1
A:Experimental source: strain IL144
C:Genetics:
A:Gene: bchH
C:Superfamily: Rhodobacter capsulatus magnesium-protoporphyrin O-methyltransferase

Query Match 85.0%; Score 34; DB 2; Length 1236;
Best Local Similarity 88.9%; Pred. No. 1.4e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPQILL 9
|||||
Db 946 LLPQIKLL 954

RESULT 14

H75201
hypothetical protein PAB0088 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 08-Sep-2000
C:Accession: H75201
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A:Reference number: A75001
A:Accession: H75201
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:95457433; PIDN:CAB49063.1; PID:el51
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0088
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0088

Query Match 82.5%; Score 33; DB 2; Length 146;
Best Local Similarity 66.7%; Pred. No. 26;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPQILL 9
|||||
Db 68 LLPQIIIV 76

RESULT 15

H84314
cytochrome aa3 controlling protein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84314
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, J.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja, J.; Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.;
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483

A:Accession: H84314
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <SPO>
A:Cross-references: GB:AE004437; NID:g10581096; PIDN:AAG19884.1; GSPDB:GN00138
C:Genetics:
A:Gene: ccp

Query Match 82.5%; Score 33; DB 2; Length 277;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 8
||||:|
Db 98 LLPLQVIL 105

RESULT 16

E70890
hypothetical protein Rv1978 - Mycobacterium tuberculosis (strain H37RV)
C:Species: Mycobacterium tuberculosis
C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C:Accession: E70890
R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, M.; Holroyd, S.; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.; Nature 393, 537-544, 1998
A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A:Reference number: A70500; MUID:98295987
A:Accession: E70890
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-282 <COL>
A:Cross-references: GB:AL022073; GB:AL123456; NID:g3256024; PIDN:CAA17851.1; PID:el25651
A:Experimental source: strain H37RV
C:Genetics:
A:Gene: Rv1978

Query Match 82.5%; Score 33; DB 2; Length 282;
Best Local Similarity 77.8%; Pred. No. 51;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
||||:|
Db 197 LLPLHLLL 205

RESULT 17

A25970
transcription activator of cholera toxin - Vibrio cholerae
C:Species: Vibrio cholerae
C:Date: 16-Aug-1988 #sequence_revision 16-Aug-1988 #text_change 08-Oct-1999
C:Accession: A25970; A38435; B39266
R:Miller, V.L.; Taylor, R.K.; Mekalanos, J.J.
Cell 48, 271-279, 1987
A:Title: Cholera toxin transcriptional activator ToxR is a transmembrane DNA binding protein
A:Reference number: A25970; MUID:87102874
A:Accession: A25970
A:Molecule type: DNA
A:Residues: 1-294 <MIL>
A:Cross-references: GB:M21249; NID:g155248; PIDN:AAA27549.1; PID:g155249
R:Dirita, V.J.; Mekalanos, J.J.
Cell 64, 29-37, 1991
A:Title: Periplasmic interaction between two membrane regulatory proteins, ToxR and ToxS
A:Reference number: A38435; MUID:91098651
A:Accession: A38435
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 263-294 <DIR>
A:Cross-references: GB:M62761; GB:M37768; NID:g155302; PIDN:AAA63558.1; PID:g155303
R:Parsot, C.; Mekalanos, J.J.
Proc. Natl. Acad. Sci. U.S.A. 87, 9898-9902, 1990

A:Title: Expression of ToxR, the transcriptional activator of the virulence factors
A:Reference number: A39266; MUID:91088618
A:Accession: B39266
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-11 <PAR>
A:Cross-references: GB:M58033
C:Genetics:
A:Gene: toxR
C:Keywords: DNA binding; transcription regulation; transmembrane protein

Query Match 82.5%; Score 33; DB 2; Length 294;
Best Local Similarity 77.8%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
||||:|
Db 190 LLPLAVLL 198

RESULT 18

E82257
cholera toxin transcription activator VC0984 [imported] - Vibrio cholerae (strain N1)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: E82257
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.; Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Seller, L. R.K.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: E82257
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-294 <HEI>
A:Cross-references: GB:AE004179; GB:AE003852; NID:g9655432; PIDN:AAF94145.1; GSPDB:G.
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0984
A:Map position: 1

Query Match 82.5%; Score 33; DB 2; Length 294;
Best Local Similarity 77.8%; Pred. No. 53;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
||||:|
Db 190 LLPLAVLL 198

RESULT 19

A46362
amyloid precursor-like protein - mouse
C:Species: Mus musculus (house mouse)
C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 24-Nov-1999
C:Accession: A46362
R:Masco, W.; Bupp, K.; Magendantz, M.; Gusella, J.F.; Tanzi, R.E.; Solomon, F.
Proc. Natl. Acad. Sci. U.S.A. 89, 10758-10762, 1992
A:Title: Identification of a mouse brain cDNA that encodes a protein related to the .
A:Reference number: A46362; MUID:93066322
A:Accession: A46362
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-653 <WAS>
A:Experimental source: brain
A:Note: sequence inconsistent with the nucleotide translation
A:Note: sequence extracted from NCBI backbone (NCBIN:118683, NCBIIP:118684)
C:Superfamily: Alzheimer's disease amyloid beta protein; animal Kunitz-type protein
C:Keywords: transmembrane protein

Query Match 82.5%; Score 33; DB 2; Length 653;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||
Db 22 LLPLSLLL 30

RESULT 20

E75440
hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C:Accession: E75440
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: E75440
A:Molecule type: DNA
A:Status: preliminary
A:Residues: 1-245 <H1>
A:Cross-references: GB:AE001958; GB:AE000513; NID:g6458805; PIDN:AAF10660.1; PID:g645881
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1083
A:Map position: 1

Query Match 80.0%; Score 32; DB 2; Length 245;
Best Local Similarity 77.8%; Pred. No. 69;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||
Db 15 LLPLALLL 23

RESULT 21

AF1535
transport protein (Truncated, N-terminal part) homolog lin0822 [imported] - Listeria inn
C:Species: Listeria innocua
C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
C:Accession: AF1535
R:Glaser, P.; Frangaul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker
.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihi, H.
D.; Jones, L.M.; Krst, U.
Science 294, 849-852, 2001
A:Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; Ma
ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland,
A:Title: Comparative genomics of Listeria species.
A:Reference number: AB1077; MUID:21537279; PMID:11679669
A:Accession: AF1535
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-253 <GLA>
A:Cross-references: GB:AL592022; PIDN:CAC96054.1; PID:gl6413273; GSPDB:GN00178
A:Experimental source: strain Clip11262
C:Genetics:
A:Gene: lin0822

Query Match 80.0%; Score 32; DB 2; Length 253;
Best Local Similarity 66.7%; Pred. No. 72;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||
Db 232 LLPLLEVL 240

RESULT 22

JQ0708
lipopolysaccharide-binding protein precursor - American cockroach

N:Alternate names: LPS-binding protein, hemolymph
C:Species: Periplaneta americana (American cockroach)
C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 12-Feb-1999
C:Accession: A39873; JQ0708; S11323; PT0075
R:Jomori, T.; Natori, S.
J. Biol. Chem. 266, 13318-13323, 1991
A:Title: Molecular cloning of cDNA for lipopolysaccharide-binding protein from the he
ute phase expression.
A:Reference number: A39873; MUID:91302364
A:Accession: A39873
A:Molecule type: mRNA
A:Residues: 1-256 <JOM>
A:Cross-references: GB:D00711; NID:g217377; PID:d1001073; PID:g217378
A:Note: part of this sequence, including the amino end of the mature protein, was con
R:Jomori, T.; Natori, S.
submitted to JIPID, October 1990
A:Reference number: JQ0708
A:Accession: JQ0708
A:Molecule type: mRNA
A:Residues: 1-256 <JOM2>
R:Jomori, T.; Kubo, T.; Natori, S.
Eur. J. Biochem. 190, 201-206, 1990
A:Title: Purification and characterization of lipopolysaccharide-binding protein from
A:Reference number: S11323; MUID:90306004
A:Accession: S11323
A:Molecule type: protein
A:Residues: 34-48 <JO2>
A:Experimental source: hemolymph
C:Comment: This hemolymph protein binds to tetrasaccharides at the proximal end of 11

Query Match 80.0%; Score 32; DB 2; Length 256;
Best Local Similarity 66.7%; Pred. No. 72;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||
Db 7 LLPLSVLL 15

RESULT 23

B43750
chloramphenicol resistance protein homolog araJ precursor - Escherichia coli
C:Species: Escherichia coli
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: B43750; D64768; S27549
R:Reeder, T.; Schleif, R.
J. Bacteriol. 173, 7765-7771, 1991
A:Title: Mapping, sequence, and apparent lack of function of araJ, a gene of the Esch
A:Reference number: A43750; MUID:92078081
A:Accession: B43750
A:Molecule type: DNA
A:Residues: 1-394 <REE>
A:Cross-references: EMBL:M64787; NID:gl45326; PIDN:AAA23474.1; PID:gl45328
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: D64768
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-394 <BLAT>
A:Cross-references: GB:AE000145; GB:U00096; NID:gl786580; PIDN:AAC73499.1; PID:gl7865
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:

A:Gene: araj
A:Map position: 9 min
C:Superfamily: Streptomyces lividans chloramphenicol resistance protein
C:Keywords: antibiotic resistance; transmembrane protein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-394/Product: araj protein #status predicted <MAT>
F:44-60/Domain: transmembrane #status predicted <TM1>
F:70-86/Domain: transmembrane #status predicted <TM2>
F:128-144/Domain: transmembrane #status predicted <TM3>
F:160-176/Domain: transmembrane #status predicted <TM4>
F:203-219/Domain: transmembrane #status predicted <TM5>
F:235-251/Domain: transmembrane #status predicted <TM6>
F:267-283/Domain: transmembrane #status predicted <TM7>
F:292-308/Domain: transmembrane #status predicted <TM8>
F:357-373/Domain: transmembrane #status predicted <TM9>

Query Match 80.0%; Score 32; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PLOIILL 9
|||||
Db 309 PLOIILL 315

RESULT 24
B85535
probable transport protein AraJ ECs0446 [imported] - Escherichia coli (strain O157:H7, S
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 27-Nov-2001
C:Accession: B85535
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: B85535
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <STO>
A:Cross-references: GB:AE005174; NID:g12513239; PIDN:AAG54742.1; GSPDB:GN00145; UWGP:204
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: araj
C:Superfamily: Streptomyces lividans chloramphenicol resistance protein

Query Match 80.0%; Score 32; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PLOIILL 9
|||||
Db 309 PLOIILL 315

RESULT 25
F90684
probable transport protein AraJ ECs0446 [imported] - Escherichia coli (strain O157:H7, S
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 27-Nov-2001
C:Accession: F90684
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: F90684
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-394 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA833869.1; PID:g13359903; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:

A:Gene: ECs0446
C:Superfamily: Streptomyces lividans chloramphenicol resistance protein

Query Match 80.0%; Score 32; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PLOIILL 9
|||||
Db 309 PLOIILL 315

RESULT 26
A57468
P-selectin glycoprotein ligand PSGL-1 precursor, long splice form - human
N:Contains: P-selectin glycoprotein ligand PSGL-1, long splice form; P-selectin glyco
C:Species: Homo sapiens (man)
C:Date: 08-Dec-1995 #sequence_revision 08-Dec-1995 #text_change 08-Oct-1999
C:Accession: A57468; A49267; A54711
R:Veldman, G.M.; Bean, K.M.; Cumming, D.A.; Eddy, R.L.; Sait, S.N.J.; Shows, T.B.
J. Biol. Chem. 270, 16470-16475, 1995
A:Title: Genomic organization and chromosomal localization of the gene encoding huma
A:Reference number: A57468; MUID:95332364
A:Accession: A57468
A:Molecule type: DNA
A:Residues: 1-412 <VEL>
A:Cross-references: GB:U25955
R:Sako, D.; Chang, X.J.; Barone, K.M.; Vachino, G.; White, H.M.; Shaw, G.; Veldman, (Cell 75, 1179-1186, 1993
A:Title: Expression cloning of a functional glycoprotein ligand for P-selectin.
A:Reference number: A49267; MUID:94084793
A:Accession: A49267
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-134,145-412 <SAK>
A:Cross-references: GB:U02297; NID:g435416; PIDN:AAC50061.1; PID:g435417
R:Moore, K.L.; Eaton, S.F.; Lyons, D.E.; Lichenstein, H.S.; Cummings, R.D.; McEver, J. Biol. Chem. 269, 23318-23327, 1994
A:Title: The P-selectin glycoprotein ligand from human neutrophils displays sialylat
A:Reference number: A54711; MUID:94365038
A:Accession: A54711
A:Molecule type: protein
A:Residues: 350-355;390-391, 'K',393-396 <MOO>
C:Genetics:
A:Gene: GDB:SELP1G
A:Cross-references: GDB:273698; OMIM:600738
A:Map position: 12q24-12q24
A:Introns: #status absent
C:Keywords: disulfide bond; glycoprotein; polymorphism; tandem repeat; transmembrane
F:1-412/Product: P-selectin glycoprotein ligand PSGL-1 precursor, long splice form #
F:1-134,145-412/Product: P-selectin glycoprotein ligand PSGL-1 precursor, short spl
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-319/Domain: extracellular #status predicted <EXT>
F:117-261/Region: 10-residue repeats (A-T/M-E-A-Q-T-T-X-P/L-A/T)
F:320-343/Domain: transmembrane #status predicted <TMW>

Query Match 80.0%; Score 32; DB 2; Length 412;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
|||||
Db 1 MPLQILL 8

RESULT 27
JC6197
stromelysin 3 (EC 3.4.24.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 17-Mar-1999
C:Accession: JC6197
R:Okada, A.; Saex, S.; Misumi, Y.; Basset, P.
Gene 185, 187-193, 1997

A:Title: Rat stromelysin 3: cDNA cloning from healing skin wound, activation by furin and
A:Reference number: JC6197; MUID:97208872
A:Contents: Skin wounds
A:Accession: JC6197
A:Molecule type: mRNA
A:Residues: 1-491 <OKA>
A:Cross-references: GB:046034
C:Comment: This protein is a member of the matrix metalloproteinase family.
C:Superfamily: Interstitial collagenase; hemopexin repeat homology; matrix metalloproteinase
C:Keywords: hydrolase; metalloproteinase; zinc; zymogen
F:52-261/Domain: matrix metalloproteinase homology <MMP>
F:291-483/Domain: hemopexin repeat homology <PXN>
F:84,218,222,228/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F:218,222,228/Binding site: zinc, catalytic (His) (active) #status predicted
F:219/Active site: Glu #status predicted

Query Match 80.0%; Score 32; DB 2; Length 491;
Best Local Similarity 77.8%; Pred. NO. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQLILL 9
|||||:||||
DB 16 LLPLPLLLL 24

RESULT 28
A44399
stromelysin 3 (EC 3.4.24.-) - mouse
N:Alternate names: matrix metalloproteinase 11 (MMP11)
C:Species: Mus musculus (house mouse)
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 10-Jul-1998
C:Accession: A44399
R:Lefebvre, O.; Wolf, C.; Limacher, J.M.; Hutin, P.; Wendling, C.; LeMeur, M.; Basset, F.
J. Cell Biol. 119, 997-1002, 1992
A:Title: The breast cancer-associated stromelysin-3 gene is expressed during mouse mammary
A:Reference number: A44399; MUID:93054930
A:Accession: A44399
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-492 <LEF>
A:Note: sequence extracted from NCBI backbone (NCBIP:117216)
C:Superfamily: Interstitial collagenase; hemopexin repeat homology; matrix metalloproteinase
C:Keywords: hydrolase; metalloproteinase; zinc; zymogen
F:52-262/Domain: matrix metalloproteinase homology <MMP>
F:292-484/Domain: hemopexin repeat homology <PXN>
F:84,219,223,229/Binding site: zinc, catalytic (Cys, His, His, His) (inhibited) #status
F:219,223,229/Binding site: zinc, catalytic (His) (active) #status predicted
F:220/Active site: Glu #status predicted

Query Match 80.0%; Score 32; DB 2; Length 492;
Best Local Similarity 77.8%; Pred. NO. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQLILL 9
|||||:||||
DB 16 LLPLPLLLL 24

RESULT 29
AI2088
Na+/H+-exchanging protein [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AI2088
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,
N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AI2088
A:Status: preliminary
A:Molecule type: DNA

A:Residues: 1-543 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAB73963.1; PID:gl7131355; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2264

Query Match 80.0%; Score 32; DB 2; Length 543;
Best Local Similarity 75.0%; Pred. NO. 1.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQLILL 8
|||||:||||
DB 194 LLPLQILML 201

RESULT 30
T29010
hypothetical protein ZK328.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T29010
R:Favella, A.
submitted to the EMBL Data Library, March 1996
A:Description: The sequence of C. elegans cosmid ZK328.
A:Reference number: Z20552
A:Accession: T29010
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1041 <FAV>
A:Cross-references: EMBL:U50193; PIDN:AAA91251.1; CESP:ZK328.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Gene: CESP:ZK328.1
A:Introns: 25/1; 64/1; 143/1; 231/2; 285/3; 487/1; 567/2; 643/1; 937/3; 987/3; 1018/3

Query Match 80.0%; Score 32; DB 2; Length 1041;
Best Local Similarity 77.8%; Pred. NO. 3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQLILL 9
|||||:||||
DB 349 LYPLEILL 357

RESULT 31
A31759
peptidyl-dipeptidase A (EC 3.4.15.1) precursor, renal and pulmonary splice form - human
N:Alternate names: angiotensin I-converting enzyme (ACE); CD143; dipeptidyl carboxypeptidase
C:Species: Homo sapiens (man)
C:Date: 07-Jun-1990 #sequence_revision 02-Jul-1998 #text_change 18-Jun-1999
C:Accession: A31759; PQ0004
R:Soubrier, F.; Alhenc-Gelas, F.; Hubert, C.; Allegrini, J.; John, M.; Tregear, G.; C.
Proc. Natl. Acad. Sci. U.S.A. 85, 9386-9390, 1988
A:Title: Two putative active centers in human angiotensin I-converting enzyme revealed
A:Reference number: A31759; MUID:89071703
A:Accession: A31759
A:Molecule type: mRNA
A:Residues: 1-1306 <SQU>
A:Cross-references: GB:J04144; NID:gl78285; PIDN:AAA51684.1; PID:gl78286
A:Experimental source: kidney
A:Note: parts of this sequence, including the amino end of the mature protein, were d
R:Takeuchi, K.; Shimizu, T.; Ohishi, N.; Seyama, Y.; Takaku, F.; Yotsumoto, H.
J. Biochem. 106, 442-445, 1989
A:Title: Purification of human lung angiotensin-converting enzyme by high-performance
A:Reference number: PQ0004; MUID:90110025
A:Accession: PQ0004
A:Molecule type: protein
A:Residues: 'XX',32-34,'E',36-37,'X',39-41,'R',43-46 <TAK>
A:Experimental source: lung
C:Comment: This splice form is found in many tissues, in particular kidney and lung v
C:Genetics:
A:Gene: GDB:DCPL; ACE
A:Cross-references: GDB:119840; OMIM:106180

A:Map position: 17q23-17q23
C:Function:
A:Description: catalyzes the hydrolysis of dipeptides from the carboxyl end of polypeptide
A:Note: plays a role in the control of blood pressure by catalyzing the conversion of an
C:Superfamily: mammalian peptidyl-dipeptidase A
C:Keywords: alternative splicing; blood pressure control; glycoprotein; kidney; lung; me
F:1-29/Domain: signal sequence #status predicted <SIG>
F:30-1306/Domain: peptidyl dipeptidase I #status predicted <MAT>
F:1260-1276/Domain: transmembrane #status predicted <VRM>
F:38-54, 74, 111, 146, 160, 318, 445, 509, 523, 677, 695, 714, 760, 942, 1191, 1225/Binding site: carb
F:390,394/Binding site: zinc (His) #status predicted
F:988,992,1008/Binding site: zinc, catalytic (His, His, Glu) #status predicted
F:989/Active site: Glu #status predicted

Query Match 80.0%; Score 32; DB 1; Length 1306;
Best Local Similarity 77.8%; Pred. No. 3.7e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | : | | |
Db 13 LLPLQLLL 21

RESULT 32
S23756
CFTR protein - African clawed frog
C:Species: Xenopus laevis (African clawed frog)
C:Date: 05-Mar-1994 #sequence_revision 01-Sep-1995 #text_change 02-Feb-2001
C:Accession: S23756
R:Tucker, S.J.; Tannahill, D.; Higgins, C.F.
Hum. Mol. Genet. 1, 77-82, 1992
A:Title: Identification and developmental expression of the Xenopus laevis cystic fibros
A:Reference number: S23756; MUID:93244789
A:Accession: S23756
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1485 <TUC>
A:Cross-references: EMBL:X65256; NID:g64622; PIDN:CA46348.1; PID:g64623
C:Superfamily: cystic fibrosis transmembrane conductance regulator; ATP-binding cassette
C:Keywords: ATP; nucleotide binding; P-loop; transmembrane protein
F:442-623/Domain: ATP-binding cassette homology <ABC1>
F:459-466/Region: nucleotide-binding motif A (P-loop)
F:1230-1422/Domain: ATP-binding cassette homology <ABC2>
F:1247-1254/Region: nucleotide-binding motif A (P-loop)

Query Match 80.0%; Score 32; DB 2; Length 1485;
Best Local Similarity 66.7%; Pred. No. 4.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | : | | |
Db 204 IAPQLVLL 212

RESULT 33
A83214
Probable transcription regulator PA3458 [imported] - Pseudomonas aeruginosa (strain PA01
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: A83214
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
-; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: A83214
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-157 <STO>
A:Cross-references: GB:AE004766; GB:AE004091; NID:g9949593; PIDN:AAG06846.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:

A:Gene: PA3458

Query Match 77.5%; Score 31; DB 2; Length 157;
Best Local Similarity 66.7%; Pred. No. 70;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | : | | |
Db 40 LTPQLLL 48

RESULT 34
T43766
hypothetical protein 189 [imported] - slime mold (Dictyostellium discoideum) mitochon
C:Species: mitochondrion Dictyostellium discoideum
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 19-May-2000
C:Accession: T43766
R:Ogawa, S.; Yoshino, R.; Angata, K.; Pl, M.; Iwamoto, M.; Kuroe, K.; Matsuo, K.; Mo
submitted to the EMBL Data Library, December 1996
A:Description: The mitochondrial DNA of Dictyostellium discoideum. Complete sequence,
A:Reference number: Z22666
A:Accession: T43766
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-189 <OGA>
A:Cross-references: EMBL:AB000109; PIDN:BAA78070.1
C:Genetics:
A:Genome: mitochondrion
C:Superfamily: Dictyostellium mitochondrion hypothetical protein 189
C:Keywords: mitochondrion

Query Match 77.5%; Score 31; DB 2; Length 189;
Best Local Similarity 75.0%; Pred. No. 84;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 8
| | | | : | | |
Db 106 LPLQLDV 113

RESULT 35
G65039
hypothetical protein b2612 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: G65039
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G65039
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-196 <BLAT>
A:Cross-references: GB:AE000347; GB:U00096; NID:g2367142; PIDN:AAC75661.1; PID:g1788
A:Experimental source: strain K-12, substrain MGL655

Query Match 77.5%; Score 31; DB 2; Length 196;
Best Local Similarity 66.7%; Pred. No. 87;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | : | | |
Db 124 LAPQLLMM 132

RESULT 36
A54984
ELF-1 protein precursor - mouse
N:Alternate names: Cek7 ligand
C:Species: Mus musculus (house mouse)
C:Date: 11-Nov-1994 #sequence_revision 11-Nov-1994 #text_change 29-Sep-1999

C:Accession: A54984; A55873
R:Cheng, H.J.; Flanagan, J.G.
Cell 79, 157-168, 1994
A:Title: Identification and cloning of ELF-1, a developmentally expressed ligand for the A:Reference number: A54984; MUID:95007776
A:Accession: A54984
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-209 <CHE>
A:Cross-references: GB:U14941; NID:9558836; PIDN:AAA53636.1; PID:g558837
R:Shao, H.; Lou, L.; Pandey, A.; Verderame, M.F.; Stever, D.A.; Dixit, V.M.
J. Biol. Chem. 270, 3467-3470, 1995
A:Title: cDNA cloning and characterization of a Cek7 receptor protein-tyrosine kinase 1 A:Reference number: A55873; MUID:95181289
A:Accession: A55873
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-209 <SHA>
A:Cross-references: GB:U14752; NID:9681886; PIDN:AAA68520.1; PID:g681887
C:Superfamily: axon guidance signal protein
C:Keywords: lipoprotein; membrane protein

Query Match 77.5%; Score 31; DB 2; Length 209;
Best Local Similarity 77.8%; Pred. No. 93;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLOIILL 9
| | | | : | | |
Db 8 LLPLELLLLL 16

RESULT 37
JE0322
eprlin-A2 - human
C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 21-Jul-2000
C:Accession: JE0322
R:Aasheim, H.; Pedetour, F.; Grosgeorge, J.; Logtenberg, T.
Biochem. Biophys. Res. Commun. 252, 378-382, 1998
A:Title: Cloning, chromosomal mapping, and tissue expression of the gene encoding the huma A:Reference number: JE0322; MUID:99043414
A:Accession: JE0322
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-213 <AAS>
A:Cross-references: GB:AJ007292; NID:93688367; PIDN:CAA07435.1; PID:g3688368
C:Superfamily: axon guidance signal protein

Query Match 77.5%; Score 31; DB 2; Length 213;
Best Local Similarity 77.8%; Pred. No. 95;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLOIILL 9
| | | | : | | |
Db 8 LLPLELLLLL 16

RESULT 38
C84914
hypothetical protein At2g47360 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84914
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: C84914
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <STO>

A:Cross-references: GB:AE002093; NID:g2275214; PIDN:AA63836.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g47360
A:Map position: 2

Query Match 77.5%; Score 31; DB 2; Length 303;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LLPLOIILL 9
| | | | : | | |
Db 31 LLPLSFLLL 39

RESULT 39
H86811
sugar ABC transporter permease protein ypcH [imported] - Lactococcus lactis subsp. la C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 31-Dec-2001
C:Accession: H86811
R:Boilotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarne, K.; Weissenbach, J.; Eh Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: H86811
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-306 <STO>
A:Cross-references: GB:AE005176; PID:gl2724492; PIDN:AAK05594.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: ypcH
C:Superfamily: maltose transport protein malC

Query Match 77.5%; Score 31; DB 2; Length 306;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLOIILL 9
| | | | : | | |
Db 233 LIPLQYLLM 241

RESULT 40
A41720
acid phosphatase (EC 3.1.3.2) 5 precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: A41720
R:Ek-Rylander, B.; Bill, P.; Norgard, M.; Nilsson, S.; Andersson, G.
J. Biol. Chem. 266, 24684-24689, 1991
A:Title: Cloning, sequence, and developmental expression of a type 5, tartrate-resist A:Reference number: A41720; MUID:92105136
A:Accession: A41720
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-327 <EKA>
A:Cross-references: GB:M76110; NID:g207543; PIDN:AAA42305.1; PID:g207544
C:Superfamily: tartrate-resistant acid phosphatase; phosphoesterase core homology C:Keywords: phosphoric monoester hydrolase
F:29-114/Domain: phosphoesterase core homology <PEC>

Query Match 77.5%; Score 31; DB 1; Length 327;
Best Local Similarity 88.9%; Pred. No. 1.5e+02;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLOIILL 9
| | | | : | | |
Db 7 LLSLQIILL 15

RESULT 41
A41927

insulin-like growth factor-binding protein 2 precursor - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1993 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
C:Accession: A41927; JQ1029; A53957; S05389; A35803; C54651
R:Binkert, C.; Margot, J.B.; Landwehr, J.; Heinrich, G.; Schwander, J.
Mol. Endocrinol. 6, 826-836, 1992
A:Title: Structure of the human insulin-like growth factor binding protein-2 gene.
A:Reference number: A41927; MUID:92293159
A:Accession: A41927
A:Molecule type: DNA
A:Residues: 1-328 <BIN>
A:Cross-references: GB:S37730; NID:g250198; PIDN:AAB22308.1; PID:g250199
A:Experimental source: placenta
R:Ehrenborg, E.; Vilhelmsdottir, S.; Bajjalica, S.; Larsson, C.; Sterm, I.; Koch, J.; Bickel, C.; Margot, J.B.; Landwehr, J.; Heinrich, G.; Schwander, J.
Biochem. Biophys. Res. Commun. 176, 1250-1255, 1991
A:Title: Structure and localization of the human insulin-like growth factor-binding protein-2.
A:Reference number: JQ1029; MUID:91248211
A:Accession: JQ1029
A:Molecule type: DNA
A:Residues: 1-59, 'R', '61-322, 'D', 324-328 <EHR>
A:Cross-references: GB:M69237
A:Note: the authors translated the codon CGG for residue 60 as Pro and the codon GAC for Arg.
R:Agarwal, N.; Hsieh, C.L.; Sills, D.; Swaroop, M.; Desai, B.; Francke, U.; Swaroop, A.
Exp. Eye Res. 52, 549-561, 1991
A:Title: Sequence analysis, expression and chromosomal localization of a gene, isolated from a human retina.
A:Reference number: A53957; MUID:91293227
A:Accession: A53957
A:Molecule type: mRNA
A:Residues: 1-328 <AGA>
A:Cross-references: GB:M35410; NID:g179476; PIDN:AAA03246.1; PID:g179477
A:Experimental source: retina
R:Binkert, C.; Landwehr, J.; Mary, J.L.; Schwander, J.; Heinrich, G.
EMBO J. 8, 2497-2502, 1989
A:Title: Cloning, sequence analysis and expression of a cDNA encoding a novel insulin-like growth factor-binding protein-2.
A:Reference number: S05389; MUID:90060007
A:Accession: S05389
A:Molecule type: mRNA
A:Residues: 1-319, 'C', 321-328 <BI2>
A:Cross-references: EMBL:X16302; NID:g33009; PIDN:CAA34373.1; PID:g33010
R:Zapp, J.; Klefer, M.; Merryweather, J.; Maslarz, F.; Bauer, D.; Born, W.; Fischer, J.A.
J. Biol. Chem. 265, 14892-14898, 1990
A:Title: Isolation from adult human serum of four insulin-like growth factor (IGF) binding proteins.
A:Reference number: A35803; MUID:90368661
A:Accession: A35803
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-19, 'P', 21-328 <ZAP>
R:Roghani, M.; Segovia, B.; Whitechurch, O.; Binoux, M.
Growth Regul. 1, 125-130, 1991
A:Title: Purification from human cerebrospinal fluid of insulin-like growth factor binding protein-2.
A:Reference number: A54651; MUID:93091816
A:Accession: C54651
A:Molecule type: protein
A:Residues: 43-44, 'X', 46-47, 'X', 49-55, 'X', 57-61 <ROG>
A:Experimental source: cerebrospinal fluid
A:Note: sequence extracted from NCB1 backbone (NCBIP:121072)
C:Genetics:
A:Gene: GDB:IGFBP2; IBP2
A:Cross-references: GDB:126735; OMIM:146731
A:Map position: 2q33-2q34
A:Introns: 151/1; 227/3; 274/3
C:Superfamily: insulin-like growth factor binding protein 1; thyroglobulin type I repeat
C:Keywords: plasma
F:1-39/Domain: signal sequence #status predicted <SIG>
F:40-328/Product: insulin-like growth factor-binding protein 2 #status predicted <MAT>
F:230-309/Domain: thyroglobulin type I repeat homology <THY1>
F:304-306/Region: cell attachment (R-G-D) motif
Query Match 77.5%; Score 31; DB 1; Length 328;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLPLQIILL 9
Db 20 LLPLQLLLL 28
RESULT 42
AB3334
daunorubicin resistance transmembrane protein [imported] - Brucella melitensis (stra
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AB3334
R:DelVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujter, C.; Los, T.; Ivano
; Mazur, M.; Goldsman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Le
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella meli
A:Reference number: AD3252; PMID:11756688
A:Accession: AB3334
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-370 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51837.1; PID:g17982584; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMIEI0656
A:Map position: I
Query Match 77.5%; Score 31; DB 2; Length 370;
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLPLQIL 7
Db 294 LLPLQVL 300
RESULT 43
A98157
probable permease of ABC transporter PA5230 [imported] - Agrobacterium tumefaciens (C
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: A98157
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Gold
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: A98157
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-372 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK88779.1; PID:g15158527; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_429
A:Map position: linear chromosome
Query Match 77.5%; Score 31; DB 2; Length 372;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLPLQIL 7
Db 296 LLPLQVL 302
RESULT 44
AH3130
ABC transporter, membrane spanning protein Atu4668 [imported] - Agrobacterium tumefa
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AH3130
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Wo

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer *Agrobacterium tumefaciens* C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AH3130

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-372 <KUR>

A:Cross-references: GB:AE008689; PIDN:AA45462.1; PID:gl7743167; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu4668

A:Map position: linear chromosome

Query Match 77.5%; Score 31; DB 2; Length 372;

Best Local Similarity 85.7%; Pred. No. 1.7e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQIL 7

|||||

Db 296 LLPLOVL 302

RESULT 45

T26360

hypothetical protein Y102A5C.21 - *Caenorhabditis elegans*

C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 04-Mar-2000

C:Accession: T26360

R:Gardner, A.

submitted to the EMBL Data Library, September 1998

A:Reference number: Z20204

A:Accession: T26360

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-388 <WIL>

A:Cross-references: EMBL:AL031627; PIDN:CAA20962.1; CESP:Y102A5C.21

A:Experimental source: clone Y102A5C

C:Genetics:

A:Gene: CESP:Y102A5C.21

A:Introns: 154/1; 364/1

C:Superfamily: *Caenorhabditis* hypothetical protein C4967.2

Query Match

Best Local Similarity 77.5%; Score 31; DB 2; Length 388;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LPLQILLL 9

|||||

Db 74 LPLQILAL 81

Search completed: November 6, 2002, 12:08:16

Job time : 13 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 8.11111 Seconds
(without alignments)
27.102 Million cell updates/sec

Title: US-09-905-083-35
Perfect score: 40
Sequence: 1 LLPLQLLLL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	40	100.0	9	4	US-09-502-600-35
2	40	100.0	253	2	US-08-557-146-2
3	40	100.0	253	2	US-08-824-874-3
4	40	100.0	253	2	US-09-154-344-2
5	40	100.0	253	3	US-08-930-188-2
6	40	100.0	253	4	US-09-210-084-3
7	40	100.0	253	5	PCT-US96-04294-2
8	36	90.0	9	4	US-09-502-600-33
9	33	82.5	190	1	US-08-339-152A-19
10	33	82.5	190	2	US-08-007-999B-6
11	33	82.5	190	2	US-08-689-276A-6
12	33	82.5	232	2	US-08-956-047-36
13	33	82.5	634	1	US-08-339-152A-17
14	33	82.5	653	1	US-08-339-152A-16
15	33	82.5	653	2	US-08-007-999B-3
16	33	82.5	653	2	US-08-689-276A-3
17	33	82.5	2787	4	US-09-245-041-15
18	32	80.0	9	4	US-09-502-600-36
19	32	80.0	12	4	US-09-502-600-28
20	32	80.0	264	4	US-08-713-556F-42
21	32	80.0	269	4	US-08-713-556F-38
22	32	80.0	313	4	US-08-713-556F-36
23	32	80.0	402	2	US-08-477-254A-2
24	32	80.0	402	2	US-08-472-576B-2
25	32	80.0	402	2	US-08-428-734B-2
26	32	80.0	402	3	US-09-063-237-1
27	32	80.0	402	4	US-08-713-556F-2

28	32	80.0	412	2	US-08-477-254A-4	Sequence 4, Appli
29	32	80.0	412	2	US-08-472-576B-4	Sequence 4, Appli
30	32	80.0	412	2	US-08-428-734B-4	Sequence 4, Appli
31	32	80.0	412	4	US-08-713-556F-4	Sequence 4, Appli
32	32	80.0	492	1	US-07-794-393-4	Sequence 4, Appli
33	32	80.0	492	1	US-08-001-711-4	Sequence 4, Appli
34	32	80.0	1306	4	US-08-989-299-7	Sequence 7, Appli
35	31	77.5	209	1	US-08-455-001-2	Sequence 2, Appli
36	31	77.5	209	4	US-08-308-814-2	Sequence 2, Appli
37	31	77.5	209	5	PCT-US95-11869-2	Sequence 2, Appli
38	31	77.5	213	1	US-09-609-324A-10	Sequence 10, Appli
39	31	77.5	213	2	US-08-920-440B-10	Sequence 10, Appli
40	31	77.5	213	4	US-09-173-492-10	Sequence 10, Appli
41	31	77.5	213	4	US-09-173-133-10	Sequence 10, Appli
42	31	77.5	480	3	US-08-923-454A-18	Sequence 18, Appli
43	31	77.5	990	4	US-09-627-376-7	Sequence 7, Appli
44	31	77.5	993	4	US-09-060-410-4	Sequence 4, Appli
45	31	77.5	1724	2	US-08-477-451-15	Sequence 15, Appli

ALIGNMENTS

RESULT 1
US-09-502-600-35
; Sequence 35, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US/09/502,600A
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 35
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 6-14 of the SCCE protein
US-09-502-600-35

Query Match 100.0%; Score 40; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	LLPLQLLLL	9	
Db	1	LLLLLLLLL	9	

RESULT 2
US-08-557-146-2
; Sequence 2, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-557-146-2

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
Db 6 LLPLQILL 14

RESULT 3
US-08-824-874-3
; Sequence 3, Application US/08824874
; Patent No. 5962300
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE: Filed Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

; LIBRARY: GenBank
; CLONE: 532504
US-08-824-874-3
Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLPLQILL 9
Db 6 LLPLQILL 14
RESULT 4
US-09-154-344-2
; Sequence 2, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-154-344-2

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
Db 6 LLPLQILL 14

RESULT 5
US-08-930-188-2
; Sequence 2, Application US/08930188
; Patent No. 6093397
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.

```

; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,188
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-930-188-2

Query Match 100.0%; Score 40; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
Db 6 LLPLQILL 14

RESULT 6
US-09-210-084-3
; Sequence 3, Application US/09210084
; Patent No. 6197511
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,084
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 532504
; US-09-210-084-3

Query Match 100.0%; Score 40; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
Db 6 LLPLQILL 14

RESULT 7
PCT-US96-04294-2
; Sequence 2, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
PCT-US96-04294-2

Query Match 100.0%; Score 40; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.6; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
| | | | |
Db 6 LLPLQIILL 14

RESULT 8

US-09-502-600-33
; Sequence 33, Application US/09502600A

; Patent No. 6294344

; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.

; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of

; FILE REFERENCE: D6223CIP-C

; CURRENT FILING DATE: 2000-02-11

; CURRENT APPLICATION NUMBER: US/09/502,600A

; PRIOR FILING DATE: 09/039,211

; NUMBER OF SEQ ID NOS: 136

; SEQ ID NO 33

; LENGTH: 9

; TYPE: PRT

; ORGANISM: Homo sapiens

; FEATURE:

; OTHER INFORMATION: Residues 5-13 of the SCCE protein

US-09-502-600-33

Query Match

Best Local Similarity 90.0%; Score 36; DB 4; Length 9;

Mismatches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQIILL 8
| | | | |
Db 2 LLPLQIILL 9

RESULT 9

US-08-339-152A-19

; Sequence 19, Application US/08339152A

; Patent No. 5643726

; GENERAL INFORMATION:

; APPLICANT: Tanzi, Rudolph E.

; APPLICANT: Kovacs, Dora M.

; TITLE OF INVENTION: Methods For Modulating Transcription

; TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter

; NUMBER OF SEQUENCES: 33

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.

; STREET: 1100 New York Ave., NW, Suite 600

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/339,152A

; FILING DATE: 10-NOV-1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Steffe, Eric K.

; REGISTRATION NUMBER: 36,688

; REFERENCE/DOCKET NUMBER: 0609.4120000

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-371-2600

; TELEFAX: 202-371-2540

; TELEX:

; INFORMATION FOR SEQ ID NO: 19:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 190 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

US-08-339-152A-19

Query Match

Best Local Similarity 82.5%; Score 33; DB 1; Length 190;

Mismatches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9
| | | | |
Db 2 LLPLSLLLL 10

RESULT 10

US-08-007-999B-6

; Sequence 6, Application US/08007999B

; Patent No. 5851787

; GENERAL INFORMATION:

; APPLICANT: Wasco, Wilma

; APPLICANT: Bupp, Keith

; APPLICANT: Megdantcz, Margaret

; APPLICANT: Tanzi, Rudolph

; APPLICANT: Solomon, Frank

; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES

; TITLE OF INVENTION: THEREOF

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.

; STREET: 1100 New York Ave., NW

; CITY: Washington

; STATE: DC

; COUNTRY: USA

; ZIP: 20005

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/007,999B

; FILING DATE: 21-JAN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/872,642

; FILING DATE: 20-APR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/930,022

; FILING DATE: 17-AUG-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Townsend, G. Kevin

; REGISTRATION NUMBER: 34,033

; REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (202)371-2571

; TELEFAX: (202)371-2540

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 190 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

US-08-007-999B-6

Query Match

Best Local Similarity 82.5%; Score 33; DB 2; Length 190;

Mismatches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQIILL 9

```
Db      2  LLPLSLLL 10
|||| :|||
RESULT 11
US-08-689-276A-6
; Sequence 6, Application US/08689276A
; Patent No. 5891991
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: JORGE A. GOLDSTEIN
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.3520003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-689-276A-6
Query Match      82.5%; Score 33; DB 2; Length 190;
Best Local Similarity 77.8%; Pred. No. 25;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      1  LLPLQILL 9
|||| :|||
Db      2  LLPLSLLL 10
RESULT 12
US-08-956-047-36
; Sequence 36, Application US/08956047
; Patent No. 5882924
; GENERAL INFORMATION:
; APPLICANT: Fritz, Hans-Joachim
; APPLICANT: Hennecke, Frank
; APPLICANT: Kolmar, Harald
; TITLE OF INVENTION:
; TITLE OF INVENTION: Methods For Modulating Transcription
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 22-OCT-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/257,669
; FILING DATE: 08-JUN-1994
; APPLICATION NUMBER: DE P 43 19 296.3
; FILING DATE: 10-JUN-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Pohlman, Sandra M.
; REGISTRATION NUMBER: P39,691
; REFERENCE/DOCKET NUMBER: 05552.1368-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 232 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-956-047-36
Query Match      82.5%; Score 33; DB 2; Length 232;
Best Local Similarity 77.8%; Pred. No. 31;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY      1  LLPLQILL 9
|||| :|||
Db      190 LLPLAVLLL 198
RESULT 13
US-08-339-152A-17
; Sequence 17, Application US/08339152A
; Patent No. 5643726
; GENERAL INFORMATION:
; APPLICANT: Tanzi, Rudolph E.
; APPLICANT: Kovacs, Dora M.
; TITLE OF INVENTION: Methods For Modulating Transcription
; TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/339,152A
FILING DATE: 10-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 0609.4120000
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
TELEX:
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 634 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-339-152A-17

Query Match 82.5%; Score 33; DB 1; Length 634;
Best Local Similarity 77.8%; Pred. No. 91;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
Db 2 LLPLSLLL 10

RESULT 14
US-08-339-152A-16
Sequence 16, Application US/08339152A
Patent No. 5643726
GENERAL INFORMATION:
APPLICANT: Tanzi, Rudolph E.
TITLE OF INVENTION: Methods For Modulating Transcription
FROM THE AMYLOID -PROTEIN PRECURSOR (APP) PROMOTER
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Ave., NW, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/339,152A
FILING DATE: 10-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 0609.4120000
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
TELEX:

INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 653 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-339-152A-16

Query Match 82.5%; Score 33; DB 1; Length 653;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
Db 22 LLPLSLLL 30

RESULT 15
US-08-007-999B-3
Sequence 3, Application US/08007999B
Patent No. 5851787
GENERAL INFORMATION:
APPLICANT: Wasco, Wilma
APPLICANT: Bupp, Keith
APPLICANT: Magendantz, Margaret
APPLICANT: Tanzi, Rudolph
APPLICANT: Solomon, Frank
TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
STREET: 1100 New York Ave., NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/007,999B
FILING DATE: 21-JAN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,642
FILING DATE: 20-APR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/930,022
FILING DATE: 17-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Townsend, G. Kevin
REGISTRATION NUMBER: 34,033
REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT
TELEPHONE: (202)371-2571
TELEFAX: (202)371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 653 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-007-999B-3

Query Match 82.5%; Score 33; DB 2; Length 653;
Best Local Similarity 77.8%; Pred. No. 94;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
Db 22 LLPLSLLL 30

RESULT 16
US-08-689-276A-3
Sequence 3, Application US/08689276A
Patent No. 5891991
GENERAL INFORMATION:
APPLICANT: Wasco, Wilma
APPLICANT: Bupp, Keith
APPLICANT: Magendantz, Margaret
APPLICANT: Tanzi, Rudolph
APPLICANT: Solomon, Frank

; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
 ; TITLE OF INVENTION: THEREOF
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
 ; STREET: 1100 New York Ave., NW
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/689,276A
 ; FILING DATE: 06-AUG-1996
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 08/007,999
 ; FILING DATE: 21-JAN-1993
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/872,642
 ; FILING DATE: 20-APR-1992
 ; APPLICATION NUMBER: US 07/930,022
 ; FILING DATE: 17-AUG-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: JORGE A. GOLDSTEIN
 ; REGISTRATION NUMBER: 29,021
 ; REFERENCE/DOCKET NUMBER: 0609.3520003
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)371-2571
 ; TELEFAX: (202)371-2540
 ; INFORMATION FOR SEQ ID NO: 3:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 653 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-689-276A-3

Query Match 82.5%; Score 33; DB 2; Length 653;
 Best Local Similarity 77.8%; Pred. No. 94;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILLL 9
 ||||:||||

Db 22 LLPLSLLLL 30

RESULT 17
 US-09-245-041-15
 ; Sequence 15, Application US/09245041
 ; Patent No. 6274339
 ; GENERAL INFORMATION:
 ; APPLICANT: Moore, K.
 ; APPLICANT: Nagle, D.
 ; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE DIAGNOSIS AND TREATMENT
 ; FILE REFERENCE: 7853-136
 ; CURRENT APPLICATION NUMBER: US/09/245,041
 ; CURRENT FILING DATE: 1999-02-05
 ; EARLIER APPLICATION NUMBER: 60/093,630
 ; EARLIER FILING DATE: 1998-07-21
 ; EARLIER APPLICATION NUMBER: 60/104,978
 ; EARLIER FILING DATE: 1998-10-20
 ; NUMBER OF SEQ ID NOS: 131
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 15
 ; LENGTH: 2787
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens

US-09-245-041-15

Query Match 82.5%; Score 33; DB 4; Length 2787;
 Best Local Similarity 87.5%; Pred. No. 4.4e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILLL 9

||:|||||

Db 2490 LPVQILLL 2497

RESULT 18

US-09-502-600-36
 ; Sequence 36, Application US/09502600A
 ; Patent No. 6294344
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
 ; FILE REFERENCE: D6223CIP-C
 ; CURRENT FILING DATE: 2000-02-11
 ; CURRENT APPLICATION NUMBER: US/09/502,600A
 ; PRIOR APPLICATION NUMBER: 09/039,211
 ; PRIOR FILING DATE: 03-14-1998
 ; NUMBER OF SEQ ID NOS: 136
 ; SEQ ID NO 36
 ; LENGTH: 9
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; FEATURE:
 ; OTHER INFORMATION: Residues 4-12 of the SCE protein
 ; US-09-502-600-36

Query Match 80.0%; Score 32; DB 4; Length 9;
 Best Local Similarity 100.0%; Pred. No. 1.7e+05;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQIL 7

|||||||

Db 3 LLPLQIL 9

RESULT 19

US-09-502-600-28
 ; Sequence 28, Application US/09502600A
 ; Patent No. 6294344
 ; GENERAL INFORMATION:
 ; APPLICANT: O'Brien, Timothy J.
 ; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
 ; FILE REFERENCE: D6223CIP-C
 ; CURRENT FILING DATE: 2000-02-11
 ; CURRENT APPLICATION NUMBER: US/09/502,600A
 ; PRIOR APPLICATION NUMBER: 09/039,211
 ; PRIOR FILING DATE: 03-14-1998
 ; NUMBER OF SEQ ID NOS: 136
 ; SEQ ID NO 28
 ; LENGTH: 12
 ; TYPE: PRT
 ; ORGANISM: Artificial sequence
 ; FEATURE:
 ; OTHER INFORMATION: a poly-lysine linked multiple Ag peptide derived from
 ; OTHER INFORMATION: SCE protein sequences
 ; US-09-502-600-28

Query Match 80.0%; Score 32; DB 4; Length 12;
 Best Local Similarity 100.0%; Pred. No. 2.1;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 PLQILLL 9

|||||||

Db 1 PLQILLL 7

RESULT 20

US-08-713-556F-42
: Sequence 42, Application US/08713556F
: Patent No. 6277975
: GENERAL INFORMATION:
: APPLICANT: Larsen, Glenn
: APPLICANT: Sako, Dianne
: APPLICANT: Chang, Xiao Jia
: APPLICANT: Veldman, Geertruida M.
: APPLICANT: Cumming, Dale
: APPLICANT: Kumar, Ravindra
: APPLICANT: Shaw, Gray
: TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LEGAL AFFAIRS
: STREET: 87 CAMBRIDGE PARK DRIVE
: CITY: CAMBRIDGE
: STATE: MA
: COUNTRY: USA
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/713,556F
: FILING DATE:
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/965,662
: FILING DATE: 23-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/112,608
: FILING DATE: 26-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/10168
: FILING DATE: 22-OCT-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/235,398
: FILING DATE: 28-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/316,305
: FILING DATE: 30-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/428,734
: FILING DATE: 25-APR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWN, SCOTT A.
: REGISTRATION NUMBER: 32,724
: REFERENCE/DOCKET NUMBER: GI 5213F
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 42:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 264 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-713-556F-42

Query Match 80.0%; Score 32; DB 4; Length 264;
Best Local Similarity 75.0%; Pred. No. 55;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLILL 9
:|||||
Db 1 MPLQLLLL 8

RESULT 21

US-08-713-556F-38
: Sequence 38, Application US/08713556F
: Patent No. 6277975
: GENERAL INFORMATION:
: APPLICANT: Larsen, Glenn
: APPLICANT: Sako, Dianne
: APPLICANT: Chang, Xiao Jia
: APPLICANT: Veldman, Geertruida M.
: APPLICANT: Cumming, Dale
: APPLICANT: Kumar, Ravindra
: APPLICANT: Shaw, Gray
: TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
: NUMBER OF SEQUENCES: 45
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: LEGAL AFFAIRS
: STREET: 87 CAMBRIDGE PARK DRIVE
: CITY: CAMBRIDGE
: STATE: MA
: COUNTRY: USA
: ZIP: 02140
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/713,556F
: FILING DATE:
: CLASSIFICATION: 536
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/965,662
: FILING DATE: 23-OCT-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/112,608
: FILING DATE: 26-AUG-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/10168
: FILING DATE: 22-OCT-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/235,398
: FILING DATE: 28-APR-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/316,305
: FILING DATE: 30-SEP-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/428,734
: FILING DATE: 25-APR-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: BROWN, SCOTT A.
: REGISTRATION NUMBER: 32,724
: REFERENCE/DOCKET NUMBER: GI 5213F
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (617) 498-8224
: TELEFAX: (617) 876-5851
: INFORMATION FOR SEQ ID NO: 38:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 269 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-713-556F-38

Query Match 80.0%; Score 32; DB 4; Length 269;
Best Local Similarity 75.0%; Pred. No. 56;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLILL 9
:|||||
Db 1 MPLQLLLL 8

RESULT 22
US-08-713-556F-36
; Sequence 36, Application US/08713556F
; Patent No. 6277975
; GENERAL INFORMATION:
; APPLICANT: Larsen, Glenn
; APPLICANT: Sako, Dianne
; APPLICANT: Chang, Xiao Jia
; APPLICANT: Veldman, Geertruida M.
; APPLICANT: Cumming, Dale
; APPLICANT: Kumar, Ravindra
; APPLICANT: Shaw, Gray
; TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
; NUMBER OF SEQUENCES: 45
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS
; STREET: 87 CAMBRIDGE PARK DRIVE
; CITY: CAMBRIDGE
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/713,556F
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,662
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,608
; FILING DATE: 26-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10168
; FILING DATE: 22-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235,398
; FILING DATE: 28-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,305
; FILING DATE: 30-SEP-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/428,734
; FILING DATE: 25-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWN, SCOTT A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: GI 5213F
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 36:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 313 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-713-556F-36
Query Match 80.0%; Score 32; DB 4; Length 313;
Best Local Similarity 75.0%; pred. No. 66;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LPLOILLL 9
:|||||
Db 1 MPLQLLLL 8

RESULT 23
US-08-477-254A-2
; Sequence 2, Application US/08477254A
; Patent No. 5827817
; GENERAL INFORMATION:
; APPLICANT: Larsen, Glenn
; APPLICANT: Sako, Dianne
; APPLICANT: Chang, Xiao Jia
; APPLICANT: Veldman, Geertruida M.
; APPLICANT: Cumming, Dale
; APPLICANT: Kumar, Ravindra
; APPLICANT: Shaw, Gray
; TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS
; STREET: 87 CAMBRIDGE PARK DRIVE
; CITY: CAMBRIDGE
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,254A
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,662
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,608
; FILING DATE: 26-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10168
; FILING DATE: 22-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235,398
; FILING DATE: 28-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,305
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWN, SCOTT A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: GI 5213B-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-477-254A-2
Query Match 80.0%; Score 32; DB 2; Length 402;
Best Local Similarity 75.0%; pred. No. 86;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LPLOILLL 9
:|||||
Db 1 MPLQLLLL 8

RESULT 24
US-08-472-576B-2
; Sequence 2, Application US/08472576B

```
; Patent No. 5840679
; GENERAL INFORMATION:
; APPLICANT: Larsen, Glenn
; APPLICANT: Sako, Dianne
; APPLICANT: Chang, Xiao Jia
; APPLICANT: Veldman, Geertruida M.
; APPLICANT: Cumming, Dale
; APPLICANT: Kumar, Ravindra
; APPLICANT: Shaw, Gray
; APPLICANT: Camphausen, Raymond
; TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS
; STREET: 87 CAMBRIDGE PARK DRIVE
; CITY: CAMBRIDGE
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,576B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,662
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,608
; FILING DATE: 26-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10168
; FILING DATE: 22-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235,398
; FILING DATE: 28-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,305
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWN, SCOTT A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: GI 5213E-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-472-576B-2

Query Match 80.0%; Score 32; DB 2; Length 402;
Best Local Similarity 75.0%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
Db 1 MPLQLLLL 8

RESULT 25
US-08-428-734B-2
; Sequence 2, Application US/08428734B
; Patent No. 5843707
; GENERAL INFORMATION:
; APPLICANT: Larsen, Glenn
```

```
; APPLICANT: Sako, Dianne
; APPLICANT: Chang, Xiao Jia
; APPLICANT: Veldman, Geertruida M.
; APPLICANT: Cumming, Dale
; APPLICANT: Kumar, Ravindra
; APPLICANT: Shaw, Gray
; APPLICANT: Camphausen, Raymond
; TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS
; STREET: 87 CAMBRIDGE PARK DRIVE
; CITY: CAMBRIDGE
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,734B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,662
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/112,608
; FILING DATE: 26-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/10168
; FILING DATE: 22-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/235,398
; FILING DATE: 28-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/316,305
; FILING DATE: 30-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWN, SCOTT A.
; REGISTRATION NUMBER: 32,724
; REFERENCE/DOCKET NUMBER: GI 5213E-PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 402 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-428-734B-2

Query Match 80.0%; Score 32; DB 2; Length 402;
Best Local Similarity 75.0%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
Db 1 MPLQLLLL 8

RESULT 26
US-09-063-237-1
; Sequence 1, Application US/09063237
; Patent No. 6124267
; GENERAL INFORMATION:
; APPLICANT: McEver, Rodger P.
; APPLICANT: Cummings, Richard D.
; TITLE OF INVENTION: O-Glycan Inhibitors of Selectin Mediated
; TITLE OF INVENTION: Inflammation Derived from PSGL-1
```

NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 West Peachtree
STREET: Street
CITY: Atlanta
STATE: Georgia
COUNTRY: US
ZIP: 30306-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/063,237
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/649,802
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Padst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRE110CIP7
TELECOMMUNICATION INFORMATION:
TELEPHONE: (404)873-8794
TELEFAX: (404)873-8795
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 402 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-063-237-1

Query Match 80.0%; Score 32; DB 3; Length 402;
Best Local Similarity 75.0%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
: : : : :
Db 1 MPLQLLL 8

RESULT 27
US-08-713-556F-2
Sequence 2, Application US/08713556F
Patent No. 6277975
GENERAL INFORMATION:
APPLICANT: Larsen, Glenn
APPLICANT: Sako, Dianne
APPLICANT: Chang, Xiao Jia
APPLICANT: Veldman, Geertruida M.
APPLICANT: Cumming, Dale
APPLICANT: Kumar, Ravindra
APPLICANT: Shaw, Gray
TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS
STREET: 87 CAMBRIDGE PARK DRIVE
CITY: CAMBRIDGE
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/713,556F
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA: US 07/965,662
FILING DATE: 23-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/112,608
FILING DATE: 26-AUG-1993
PRIOR APPLICATION DATA: PCT/US93/10168
FILING DATE: 22-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/235,398
FILING DATE: 28-APR-1994
PRIOR APPLICATION DATA: US 08/316,305
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/428,734
FILING DATE: 25-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWN, SCOTT A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: GI 5213F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 402 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-713-556F-2

Query Match 80.0%; Score 32; DB 4; Length 402;
Best Local Similarity 75.0%; Pred. No. 86;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
: : : : :
Db 1 MPLQLLL 8

RESULT 28
US-08-477-254A-4
Sequence 4, Application US/08477254A
Patent No. 5827817
GENERAL INFORMATION:
APPLICANT: Larsen, Glenn
APPLICANT: Sako, Dianne
APPLICANT: Chang, Xiao Jia
APPLICANT: Veldman, Geertruida M.
APPLICANT: Cumming, Dale
APPLICANT: Kumar, Ravindra
APPLICANT: Shaw, Gray
TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
NUMBER OF SEQUENCES: 34
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS
STREET: 87 CAMBRIDGE PARK DRIVE
CITY: CAMBRIDGE
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,254A

;; FILING DATE: 23-OCT-1992
;; CLASSIFICATION: 536
;; PRIOR APPLICATION DATA: US 07/965,662
;; APPLICATION NUMBER: US 07/965,662
;; FILING DATE: 23-OCT-1992
;; PRIOR APPLICATION DATA: US 08/112,608
;; APPLICATION NUMBER: US 08/112,608
;; FILING DATE: 26-AUG-1993
;; PRIOR APPLICATION DATA: PCT/US93/10168
;; APPLICATION NUMBER: PCT/US93/10168
;; FILING DATE: 22-OCT-1993
;; FILING DATE: 28-APR-1994
;; PRIOR APPLICATION DATA: US 08/235,398
;; APPLICATION NUMBER: US 08/235,398
;; FILING DATE: 28-APR-1994
;; PRIOR APPLICATION DATA: US 08/316,305
;; APPLICATION NUMBER: US 08/316,305
;; FILING DATE: 30-SEP-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BROWN, SCOTT A.
;; REGISTRATION NUMBER: 32,724
;; REFERENCE/DOCKET NUMBER: GI 5213E-PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 498-8224
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 412 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-477-254A-4

Query Match 80.0%; Score 32; DB 2; Length 412;
Best Local Similarity 75.0%; Pred. No. 89;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
:|||||
Db 1 MPLQLLL 8

RESULT 29
US-08-472-576B-4
; Sequence 4, Application US/08472576B
; Patent No. 5840679
; GENERAL INFORMATION:
; APPLICANT: Larsen, Glenn
; APPLICANT: Sako, Dianne
; APPLICANT: Chang, Xiao Jia
; APPLICANT: Veldman, Geertruida M.
; APPLICANT: Cumming, Dale
; APPLICANT: Kumar, Ravindra
; APPLICANT: Shaw, Gray
; APPLICANT: Camphausen, Raymond
; TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS
; STREET: 87 CAMBRIDGE PARK DRIVE
; CITY: CAMBRIDGE
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,576B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

;; APPLICATION NUMBER: US 07/965,662
;; FILING DATE: 23-OCT-1992
;; PRIOR APPLICATION DATA: US 08/112,608
;; APPLICATION NUMBER: US 08/112,608
;; FILING DATE: 26-AUG-1993
;; PRIOR APPLICATION DATA: PCT/US93/10168
;; APPLICATION NUMBER: PCT/US93/10168
;; FILING DATE: 22-OCT-1993
;; FILING DATE: 28-APR-1994
;; PRIOR APPLICATION DATA: US 08/235,398
;; APPLICATION NUMBER: US 08/235,398
;; FILING DATE: 28-APR-1994
;; PRIOR APPLICATION DATA: US 08/316,305
;; APPLICATION NUMBER: US 08/316,305
;; FILING DATE: 30-SEP-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: BROWN, SCOTT A.
;; REGISTRATION NUMBER: 32,724
;; REFERENCE/DOCKET NUMBER: GI 5213E-PCT
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (617) 498-8224
;; TELEFAX: (617) 876-5851
;; INFORMATION FOR SEQ ID NO: 4:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 412 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-472-576B-4

Query Match 80.0%; Score 32; DB 2; Length 412;
Best Local Similarity 75.0%; Pred. No. 89;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
:|||||
Db 1 MPLQLLL 8

RESULT 30
US-08-428-734B-4
; Sequence 4, Application US/08428734B
; Patent No. 5843707
; GENERAL INFORMATION:
; APPLICANT: Larsen, Glenn
; APPLICANT: Sako, Dianne
; APPLICANT: Chang, Xiao Jia
; APPLICANT: Veldman, Geertruida M.
; APPLICANT: Cumming, Dale
; APPLICANT: Kumar, Ravindra
; APPLICANT: Shaw, Gray
; APPLICANT: Camphausen, Raymond
; TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
; NUMBER OF SEQUENCES: 34
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEGAL AFFAIRS
; STREET: 87 CAMBRIDGE PARK DRIVE
; CITY: CAMBRIDGE
; STATE: MA
; COUNTRY: USA
; ZIP: 02140
; COMPUTER READABLE FORM:
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/428,734B
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/965,662
; FILING DATE: 23-OCT-1992
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/112,608
FILING DATE: 26-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10168
FILING DATE: 22-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/235,398
FILING DATE: 28-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,305
FILING DATE: 30-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: BROWN, SCOTT A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: GI 5213B-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-428-734B-4

Query Match 80.0% Score 32; DB 2; Length 412;
Best Local Similarity 75.0% Pred. No. 89;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
:|||||
Db 1 MPLQLLL 8

RESULT 31
US-08-713-556F-4
Sequence 4, Application US/08713556F
Patent No. 6277975
GENERAL INFORMATION:
APPLICANT: Larsen, Glenn
APPLICANT: Sako, Dianne
APPLICANT: Chang, Xiao Jia
APPLICANT: Veldman, Geertruida M.
APPLICANT: Cumming, Dale
APPLICANT: Kumar, Ravindra
APPLICANT: Shaw, Gray
TITLE OF INVENTION: NOVEL P-SELECTIN LIGAND PROTEIN
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: LEGAL AFFAIRS
STREET: 87 CAMBRIDGE PARK DRIVE
CITY: CAMBRIDGE
STATE: MA
COUNTRY: USA
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/713,556F
FILING DATE:
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/965,662
FILING DATE: 23-OCT-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/112,608
FILING DATE: 26-AUG-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10168

FILING DATE: 22-OCT-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/235,398
FILING DATE: 28-APR-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/316,305
FILING DATE: 30-SEP-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/428,734
FILING DATE: 25-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: BROWN, SCOTT A.
REGISTRATION NUMBER: 32,724
REFERENCE/DOCKET NUMBER: GI 5213F
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8224
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 412 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-713-556F-4

Query Match 80.0% Score 32; DB 4; Length 412;
Best Local Similarity 75.0% Pred. No. 89;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
:|||||
Db 1 MPLQLLL 8

RESULT 32
US-07-794-393-4
Sequence 4, Application US/07794393
Patent No. 5236844
GENERAL INFORMATION:
APPLICANT: CHAMON, PIERRE
APPLICANT: BASSET, PAUL
APPLICANT: BELLOCQ, JEAN-PIERRE
TITLE OF INVENTION: ANALYTICAL MARKERS FOR MALIGNANT BREAST
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox
STREET: 1225 Connecticut Ave. NW Suite 300
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/794,393
FILING DATE: 19911121
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9025326.1
FILING DATE: 21-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: GOLDSTEIN, JORGE A
REGISTRATION NUMBER: 29,021
REFERENCE/DOCKET NUMBER: 1383.0040000
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 466-0800
TELEFAX: (202) 833-8716
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:

```
; LENGTH: 492 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-794-393-4

Query Match      80.0%; Score 32; DB 1; Length 492;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQIILL 9
Db 16 LLPLPLLLL 24

RESULT 33
US-08-001-711-4
; Sequence 4, Application US/08001711
; Patent No. 5484726
; GENERAL INFORMATION:
; APPLICANT: BASSET, PAUL
; APPLICANT: BELLOCO, JEAN-PIERRE
; APPLICANT: CHAMON, PIERRE
; TITLE OF INVENTION: ANALYTICAL MARKERS FOR MALIGNANT BREAST
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Suite 300
; CITY: Washington
; STATE: D.C.
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/001.711
; FILING DATE: 19930107
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/794,393
; FILING DATE: 11-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9025626.1
; FILING DATE: 21-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: MILLMAN, ROBERT A
; REGISTRATION NUMBER: 36,217
; REFERENCE/DOCKET NUMBER: 1383.0040001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)466-0800
; TELEFAX: (202)833-8716
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 492 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-001-711-4

Query Match      80.0%; Score 32; DB 1; Length 492;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQIILL 9
Db 16 LLPLPLLLL 24

RESULT 34
US-08-989-299-7
; Sequence 7, Application US/08989299
; Patent No. 6194556
; GENERAL INFORMATION:
; APPLICANT: Acton, Susan L.
; APPLICANT: Robinson, Keith E.
; TITLE OF INVENTION: ANGIOTENSIN CONVERTING ENZYME HOMOLOG
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY, HOAG & ELIOT LLP
; STREET: One Post Office Square
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109-2170
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/989,299
; FILING DATE: 11-DEC-1997
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Arnold E., Beth
; REGISTRATION NUMBER: 35,430
; REFERENCE/DOCKET NUMBER: MIA-025.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-832-1000
; TELEFAX: 617-832-7000
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1306 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-989-299-7

Query Match      80.0%; Score 32; DB 4; Length 1306;
Best Local Similarity 77.8%; Pred. No. 3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQIILL 9
Db 13 LLPLPLLLL 21

RESULT 35
US-08-455-001-2
; Sequence 2, Application US/08455001
; Patent No. 5795734
; GENERAL INFORMATION:
; APPLICANT: Flanagan, John G.
; APPLICANT: Cheng, Hwai-Jong
; TITLE OF INVENTION: EPH Receptor Ligands, and Uses Related
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII (text)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/455,001
```

; FILING DATE: 31 MAY 1995
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMI-011CP2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-455-001-2

Query Match 77.5%; Score 31; DB 1; Length 209;
Best Local Similarity 77.8%; Pred. No. 67;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||

Db 8 LLPLLLLL 16

RESULT 36

US-08-308-814-2
; Sequence 2, Application US/08308814
; Patent No. 6268476

; GENERAL INFORMATION:
; APPLICANT: Flanagan, John G.
; APPLICANT: Cheng, Hwai-Jong
; TITLE OF INVENTION: EPH Receptor Ligands, and Uses Related
; TITLE OF INVENTION: Thereto
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02109

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: ASCII(txt)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/308,814
; FILING DATE: 19-SEP-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMI-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-308-814-2

Query Match 77.5%; Score 31; DB 4; Length 209;
Best Local Similarity 77.8%; Pred. No. 67;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||

Db 8 LLPLLLLL 16

RESULT 37

PCT-US95-11869-2
; Sequence 2, Application PC/TUS9511869
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: EPH Receptor Ligands, and Uses Related
; TITLE OF INVENTION: Thereto
; NUMBER OF SEQUENCES: 5
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/11869
; FILING DATE: 19-SEP-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Vincent, Matthew P.
; REGISTRATION NUMBER: 36,709
; REFERENCE/DOCKET NUMBER: HMI-011CPPC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 227-7400
; TELEFAX: (617) 227-5941
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 209 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US95-11869-2

Query Match 77.5%; Score 31; DB 5; Length 209;

Best Local Similarity 77.8%; Pred. No. 67;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||

Db 8 LLPLLLLL 16

RESULT 38

US-09-609-324A-10
; Sequence 10, Application US/09609324A
; Patent No. RE37582
; GENERAL INFORMATION:
; APPLICANT: CERRETTI, Douglas P.
; TITLE OF INVENTION: CYTOKINE DESIGNATED LERK-6
; FILE REFERENCE: A7772
; CURRENT APPLICATION NUMBER: US/09/609,324A
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 08/920,440
; PRIOR FILING DATE: 1997-08-29
; PRIOR APPLICATION NUMBER: 08/538,709
; PRIOR FILING DATE: 1995-10-03
; PRIOR APPLICATION NUMBER: 08/318,393
; PRIOR FILING DATE: 1994-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 213
; TYPE: PRT
; ORGANISM: LERK-6
US-09-609-324A-10

Query Match 77.5%; Score 31; DB 1; Length 213;
Best Local Similarity 77.8%; Pred. No. 68;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||

Db 8 LLPLLLLL 16

RESULT 39

US-08-920-440B-10

; Sequence 10, Application US/08920440B

; Patent No. 5919905

; GENERAL INFORMATION:

; APPLICANT: Cerretti, Douglas P.

; TITLE OF INVENTION: Cytokine Designated LERK-6

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Immunex Corporation.

; STREET: 51 University Street

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Apple Power Macintosh

; OPERATING SYSTEM: System 7.6

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/920,440B

; FILING DATE: 29-AUG-1997

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Henry, Janis C.

; REGISTRATION NUMBER: 34,347

; REFERENCE/DOCKET NUMBER: 2826-B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 587-0430

; TELEFAX: (206) 233-0644

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 213 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-920-440B-10

Query Match 77.5%; Score 31; DB 2; Length 213;

Best Local Similarity 77.8%; Pred. No. 68;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPQLILL 9

Db 8 LLPLLLLL 16

RESULT 40

US-09-173-492-10

; Sequence 10, Application US/09173492

; Patent No. 6194172

; GENERAL INFORMATION:

; APPLICANT: Cerretti, Douglas P.

; TITLE OF INVENTION: Cytokine Designated LERK-6

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Immunex Corporation

; STREET: 51 University Street

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Apple Power Macintosh

; OPERATING SYSTEM: System 7.6

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/173,492

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/920,440
; FILING DATE: 29-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Henry, Janis C.
; REGISTRATION NUMBER: 34,347
; REFERENCE/DOCKET NUMBER: 2826-B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 587-0430
; TELEFAX: (206) 233-0644
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 213 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-173-492-10

Query Match 77.5%; Score 31; DB 4; Length 213;

Best Local Similarity 77.8%; Pred. No. 68;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPQLILL 9

Db 8 LLPLLLLL 16

RESULT 41

US-09-173-133-10

; Sequence 10, Application US/09173133

; Patent No. 6232447

; GENERAL INFORMATION:

; APPLICANT: Cerretti, Douglas P.

; TITLE OF INVENTION: Cytokine Designated LERK-6

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Immunex Corporation

; STREET: 51 University Street

; CITY: Seattle

; STATE: WA

; COUNTRY: USA

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: Apple Power Macintosh

; OPERATING SYSTEM: System 7.6

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/173,133

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/920,440

; FILING DATE: 29-AUG-1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Henry, Janis C.

; REGISTRATION NUMBER: 34,347

; REFERENCE/DOCKET NUMBER: 2826-B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (206) 587-0430

; TELEFAX: (206) 233-0644

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 213 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-173-133-10

Query Match

Best Local Similarity

77.5%; Score 31; DB 4; Length 213;

77.8%; Pred. No. 68;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||
Db 8 LLPLLLLL 16

RESULT 42

US-08-923-454A-18
; Sequence 18, Application US/08923454A
; Patent No. 6004794
; GENERAL INFORMATION:
; APPLICANT: Creasy, Caretha
; APPLICANT: Livi, George
; APPLICANT: Karran, Eric
; APPLICANT: Clinkenbeard, Helen
; APPLICANT: Browne, Michael
; APPLICANT: Southan, Christopher
; TITLE OF INVENTION: HUMAN SERINE PROTEASE
; NUMBER OF SEQUENCES: 40
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/923,454A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/025436
; FILING DATE: 06-SEPT-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Baumeister, Kirk
; REGISTRATION NUMBER: 33,833
; REFERENCE/DOCKET NUMBER: P50547
; TELEPHONE: 610-270-5096
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 480 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
; ORIGINAL SOURCE:
; ORIGINAL SOURCE: Feature - 213 Gly/val polymorph
US-08-923-454A-18

Query Match 77.5%; Score 31; DB 3; Length 480;
Best Local Similarity 77.8%; Pred. No. 1.6e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||
Db 8 LLPLLLLL 16

RESULT 43

US-09-627-376-7
; Sequence 7, Application US/09627376

; Patent No. 6342385
; GENERAL INFORMATION:
; APPLICANT: Oi, Fengxia
; TITLE OF INVENTION: MUFACIN I BIOSYNTHESIS GENES AND PROTEINS
; FILE REFERENCE: JAB-17402/22
; CURRENT APPLICATION NUMBER: US/09/627,376
; CURRENT FILING DATE: 2001-05-30
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 7
; LENGTH: 990
; TYPE: PRT
; ORGANISM: Streptococcus mutans
US-09-627-376-7

Query Match 77.5%; Score 31; DB 4; Length 990;
Best Local Similarity 75.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLPLQILL 8
||||:||||
Db 357 LVPLQLLL 364

RESULT 44

US-09-060-410-4
; Sequence 4, Application US/09060410
; Patent No. 6165461
; GENERAL INFORMATION:
; APPLICANT: Cobb, Melanie
; APPLICANT: Hutchinson, Michele
; APPLICANT: Chen, Zhu
; APPLICANT: Berman, Kevin
; TITLE OF INVENTION: TAO PROTEIN KINASES AND METHODS OF USE
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/060,410
; FILING DATE: 14-APR-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Maki, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 860098.421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 993 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-060-410-4

Query Match 77.5%; Score 31; DB 4; Length 993;
Best Local Similarity 77.8%; Pred. No. 3.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
||||:||||

Db 967 LPLLLLLL 975

RESULT 45

US-08-477-451-15
 ; Sequence 15, Application US/08477451
 ; Patent No. 5928865
 ; GENERAL INFORMATION:
 ; APPLICANT: Covacci, Antonello
 ; TITLE OF INVENTION: Helicobacter Pylori Cagi Region
 ; NUMBER OF SEQUENCES: 46
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Chiron Corporation
 ; STREET: 4560 Horton Street
 ; CITY: Emeryville
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94608-2916
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/477,451
 ; FILING DATE: 07-JUN-1995
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: McClung, Barbara G.
 ; REGISTRATION NUMBER: 33,113
 ; REFERENCE/DOCKET NUMBER: 0335.002
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 510-601-2708
 ; TELEFAX: 510-655-3542
 ; INFORMATION FOR SEQ ID NO: 15:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1724 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-477-451-15

Query Match 77.5%; Score 31; DB 2; Length 1724;
 Best Local Similarity 66.7%; Pred. No. 6.3e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPLQLILL 9

Db 426 LPLQLILL 434

Search completed: November 6, 2002, 12:06:27
 Job time : 10.1111 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 23.7778 Seconds
(without alignments)
42.042 Million cell updates/sec

Title: US-09-905-083-35

Perfect score: 40

Sequence: 1 LLPLQLLLL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802.*

1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	40	100.0	9	22	AAE08240 Human stratum corn
2	40	100.0	136	22	ABG23378 Novel human diagno
3	40	100.0	253	16	AAR67888 Human stratum corn
4	40	100.0	253	17	AAW05383 Human amyloid prec
5	40	100.0	257	21	AAW21326 Human HSCCE. Homo
6	38	95.0	142	22	AAW63580 Human gastric canc
7	38	95.0	156	22	AAW63578 Human gastric canc
8	38	95.0	159	22	AAW63582 Human gastric canc
9	36	90.0	9	22	AAE08238 Human stratum corn
10	35	87.5	23	22	ABW43858 Peptide #11364 enc
11	35	87.5	23	22	AAW64837 Human brain expres

12	35	87.5	23	22	AAW37771	Peptide #11808 enc
13	35	87.5	480	19	AAW98431	H. pylori GHPO 446
14	34	85.0	70	20	AAV36291	Human secreted pro
15	34	85.0	73	20	AAW89030	Polypeptide fragme
16	34	85.0	73	22	ABW11130	Human secreted pro
17	34	85.0	86	22	AAO08397	Human polypeptide
18	34	85.0	224	22	AAW93417	Human protein sequ
19	34	85.0	267	18	AAW22303	Rat CRTI. Rattus
20	34	85.0	301	22	AAW87732	Human T2R02 amino
21	34	85.0	371	16	AAW75642	Bovine conglutinin
22	34	85.0	424	20	AAW89026	Polypeptide fragme
23	34	85.0	424	22	ABW11126	Human secreted pro
24	34	85.0	1001	22	AAW00784	Human bone marrow
25	34	85.0	1080	22	ABG09527	Novel human diagno
26	34	85.0	1080	22	AAW00897	Human bone marrow
27	34	85.0	1092	21	AAW42131	Human ORFX ORF1895
28	34	85.0	1092	22	AAW39421	Human polypeptide
29	34	85.0	1139	22	AAW00996	Human bone marrow
30	34	85.0	1153	22	AAW00997	Human bone marrow
31	34	85.0	1214	22	AAW41207	Human polypeptide
32	33	82.5	60	22	AAU47328	Propionibacterium
33	33	82.5	61	22	AAO12472	Human polypeptide
34	33	82.5	77	21	AAW23633	Human secreted pro
35	33	82.5	146	22	AAU14538	Human novel protei
36	33	82.5	149	22	ABW69455	Drosophila melanoq
37	33	82.5	181	22	AAU14302	Human novel protei
38	33	82.5	190	17	AAW98922	Murine APLP1 (aa21
39	33	82.5	201	21	AAW42852	Human ORFX ORF2616
40	33	82.5	201	21	AAW24060	Human PRO335 prote
41	33	82.5	201	21	AAW23604	Human secreted pro
42	33	82.5	201	21	AAW66651	Membrane-bound pro
43	33	82.5	201	22	AAW65174	Human PRO535 (UNO3
44	33	82.5	201	22	AAW50984	Human PRO335 prote
45	33	82.5	201	22	AAW53092	Human angiogenesis

ALIGNMENTS

RESULT 1

AAE08240

ID AAE08240 standard; peptide; 9 AA.

XX AC AAE08240;

XX DT 01-NOV-2001 (first entry)

XX DT Human stratum corneum chymotrypsin enzyme peptide #5 (residues 6-14).

XX DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;

XX KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;

XX KW antisense therapy; malignant hyperplasia.

XX OS Homo sapiens.

XX PN WO200159158-A1.

XX PD 16-AUG-2001.

XX PF 07-FEB-2001; 2001WO-US03977.

XX PR 11-FEB-2000; 2000US-0502600.

XX PA (UYAR-) UNIV ARKANSAS.

XX PI O'Brien TJ;

XX DR WPI; 2001-514676/56.

XX PT Diagnosing cancer comprises detecting stratum corneum chymotrypsin

XX PT enzyme -

XX XX Peptide #11364 enc

XX PS Claim 25; Page 103; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.

SQ Sequence 9 AA;
 Query Match 100.0%; Score 40; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. NO. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
 Db 1 LLPLQILL 9
 |||||

RESULT 2
 ABG23378
 ID ABG23378 standard; Protein; 136 AA.
 XX AC
 XX ABG23378;
 XX DT 18-FEB-2002 (first entry)
 XX DE Novel human diagnostic protein #23369.
 XX KW Human: chromosome mapping; gene mapping; gene therapy; forensic;
 XX KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX OS Homo sapiens.
 XX PN WO200175067-A2.
 XX PD 11-OCT-2001.
 XX PF 30-MAR-2001; 2001WO-US08631.
 XX PR 31-MAR-2000; 2000US-0540217.
 XX PR 23-AUG-2000; 2000US-0649167.
 XX PA (HYSE-) HYSEQ INC.
 XX PI Drmanac RT, Liu C, Tang YT;
 XX PI WPI; 2001-639362/73.
 XX DR N-PSDB; AAS87565.
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity.
 XX Claim 20; SEQ ID No 53737; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 136 AA;
 Query Match 100.0%; Score 40; DB 22; Length 136;
 Best Local Similarity 100.0%; Pred. NO. 4.5;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
 Db 12 LLPLQILL 20
 |||||

RESULT 3
 AAR67888
 ID AAR67888 standard; Protein; 253 AA.
 XX AC
 XX AAR67888;
 XX DT 09-AUG-1995 (first entry)
 XX DE Human stratum corneum chymotrophic recombinant enzyme (SCCE).
 XX KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
 XX KW callosities; keratosis pilaris; ichthyoses; eczema.
 XX OS Homo sapiens.
 XX PN WO9500651-A.
 XX PD 05-JAN-1995.
 XX PF 20-JUN-1994; 94WO-IB00166.
 XX PR 18-JUN-1993; 93DK-0000725.
 XX PA (SYMB-) SYMBICOM AB.
 XX PI Egelrud T, Hansson L;
 XX DR WPI; 1995-052088/07.
 XX DR N-PSDB; AAQ81203.
 XX Nucleotide sequences encoding stratum corneum chymotryptic enzyme
 PT - and related vectors, transformed cells and polypeptides, and
 PT useful for treating skin disorders, e.g. acne or psoriasis, and
 PT for identification of specific inhibitors.
 XX Disclosure; Page 97; 137pp; English.

XX The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
 CC and skin care products, especially to treat and prevent acne,
 CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
 CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
 CC produced recombinantly following mammal, insect, plant, or
 CC microorganism transformation with plasmid pS507.

SQ Sequence 253 AA;
 Query Match 100.0%; Score 40; DB 16; Length 253;
 Best Local Similarity 100.0%; Pred. NO. 8.4;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
 Db 6 LLPLQILL 14
 |||||

```

RESULT 4
AAW05383
ID AAW05383 standard; Protein: 253 AA.
XX
XX AAW05383;
XX
XX 31-DEC-1996 (first entry)
XX
XX Human amyloid precursor protein protease.
DE
XX
XX Amyloid precursor protein protease; Alzheimer's disease; diagnosis;
KW therapy.
XX
XX Homo sapiens.
OS
XX
XX WO9631122-A1.
PN
XX
XX 10-OCT-1996.
PD
XX
XX 02-APR-1996; 96WO-US04294.
PF
XX
XX 04-APR-1995; 95US-0416257.
PR
XX
XX (ELIL ) LILLY & CO ELI.
PA
XX
XX Dixon EP, Johnstone EM, Little SP;
PI
XX
XX WPI; 1996-464694/46.
DR
XX
XX N-PSDB; AAT39783.
DR
XX
XX New isolated human amyloid precursor protein protease - used to
PT develop prods. for the treatment or diagnosis of associated
PT conditions, esp. Alzheimer's disease
PT
XX
XX Claim 1; Page 44-45; 55pp; English.
PS
XX
XX Human amyloid precursor protein protease (AAW05383) is involved in
CC the processing or clearance of amyloid precursor protein to form
CC beta-amyloid peptide. Its amino acid sequence was deduced from
CC a cDNA clone (AAT39783) obtd. from a human lung library. Recombinant
CC protease can be produced in transformed or transfected prokaryotic
CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
CC used to develop products for the design and testing of cpds. useful
CC for treating or preventing conditions associated with beta-amyloid
CC peptide, esp. Alzheimer's disease.
CC
XX
XX Sequence 253 AA;
SQ
Query Match 100.0%; Score 40; DB 17; Length 253;
Best Local Similarity 100.0%; Pred. No. 8.4;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLPLQILL 9
DB 6 LLPLQILL 14
RESULT 5
AAB21326
ID AAB21326 standard; Protein: 257 AA.
XX
XX AAB21326;
XX
XX 02-FEB-2001 (first entry)
DT
XX
XX Human HSCEE.
DE
XX
XX Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCEE;
KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;
KW serine protease; cytostatic; cancer; prostrate cancer.
XX

```

```

OS Homo sapiens.
XX
XX WO200053776-A2.
XX
XX 14-SEP-2000.
PD
XX
XX 09-MAR-2000; 2000WO-CA00258.
PF
XX
XX 11-MAR-1999; 99US-0124260.
PR
XX
XX 01-APR-1999; 99US-0127386.
PR
XX
XX 21-JUL-1999; 99US-0144919.
PR
XX
XX (MOUN ) MOUNT SINAI HOSPITAL.
PA
XX
XX Yousef GM, Diamandis EP;
PI
XX
XX WPI; 2000-587440/55.
DR
XX
XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
PT protein mediated disorders, especially cancer.
PT
XX
XX Example 4; Fig 17; 184pp; English.
PS
XX
XX The present sequence is human stratum corneum chymotryptic enzyme
CC (HSCEE), a member of the kallikrein multi-gene family. Kallikreins and
CC kallikrein-like proteins are a subgroup of the serine protease enzyme
CC family. They catalyze the selective cleavage of specific polypeptide
CC precursors to release peptides with potent biological activity. Nucleic
CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
CC treatment, monitoring and diagnosis of cancers, especially prostate
CC cancer. They can also be used to identify a substance that can associate
CC with or mediate the biological activity of the proteins. Antibodies can
CC be used to treat conditions mediated by the kallikrein-like proteins.
CC
XX
XX Sequence 257 AA;
SQ
Query Match 100.0%; Score 40; DB 21; Length 257;
Best Local Similarity 100.0%; Pred. No. 8.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LLPLQILL 9
DB 6 LLPLQILL 14
RESULT 6
AAB63580
ID AAB63580 standard; Protein: 142 AA.
XX
XX AAB63580;
XX
XX 26-MAR-2001 (first entry)
DT
XX
XX Human gastric cancer associated antigen protein sequence SEQ ID NO:942.
DE
XX
XX Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
KW
XX
XX Homo sapiens.
OS
XX
XX WO200073801-A2.
PN
XX
XX 07-DEC-2000.
PD
XX
XX 26-MAY-2000; 2000WO-US14749.
PF
XX
XX 28-MAY-1999; 99US-0136526.
PR
XX
XX 10-SEP-1999; 99US-0153454.
PR
XX
XX (LUDW-) LUDWIG INST CANCER RES.
PA
XX
XX Obata Y;
PI

```

```

XX WPI; 2001-025274/03.
XX
XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
XX Example 1; Page 625; 799pp; English.
XX
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63322 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nuclei
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein
CC e.g. cancer.
XX
XX Sequence 142 AA;
XX
Query Match . 95.0%; Score 38; DB 22; Length 142;
Best Local Similarity 88.9%; Pred. NO. 11;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0
QY 1 LLPLQLLL 9
Db 48 LLPLQLLLL 56
|||||:||||
RESULT 7
AAB63578
ID AAB63578 standard; Protein; 156 AA.
XX AC AAB63578;
XX DT 26-MAR-2001 (first entry)
XX DE Human gastric cancer associated antigen protein sequence SEQ ID NO:940.
XX KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
XX OS cancer associated antigen; cytotstatic; cancer vaccine.
XX PA Homo sapiens.
XX PN WO200073801-A2.
XX PD 07-DEC-2000.
XX PF 26-MAY-2000; 2000WO-US14749.
XX PR 28-MAY-1999; 99US-0136526.
XX PR 10-SEP-1999; 99US-0153454.
XX PA (LUDW-) LUDWIG INST CANCER RES.
XX PI Obata Y;
XX DR WPI; 2001-025274/03.
XX
XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
XX Example 1; Page 624; 799pp; English.
XX
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63322 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nuclei
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein
CC e.g. cancer.
XX
XX Sequence 142 AA;
XX
Query Match . 95.0%; Score 38; DB 22; Length 142;
Best Local Similarity 88.9%; Pred. NO. 11;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0
QY 1 LLPLQLLL 9
Db 48 LLPLQLLLL 56
|||||:||||

```

```

Db      48 LLPLQLLL 56
|||||:||||
RESULT 9
AAE08238
ID      AAE08238 standard; peptide; 9 AA.
XX
AC      AAE08238;
XX
DT      01-NOV-2001 (first entry)
XX
DE      Human stratum corneum chymotrypsin enzyme peptide #3 (residues 5-13).
XX
KW      Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW      cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW      antisense therapy; malignant hyperplasia.
XX
OS      Homo sapiens.
XX
PN      WO200159158-A1.
XX
PD      16-AUG-2001.
XX
PF      07-FEB-2001; 2001WO-US03977.
XX
PR      11-FEB-2000; 2000US-0502600.
XX
PA      (UYAR-) UNIV ARKANSAS.
XX
PI      O'Brien TJ;
XX
DR      WPI; 2001-514676/56.
XX
PT      Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT      enzyme -
XX
PS      Claim 25; Page 102; 127pp; English.
XX
CC      The invention relates to diagnosing cancer especially ovarian cancer, by
CC      screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC      considered to be an integral part of tumour growth and metastasis, and
CC      therefore, markers indicative of their presence or absence are useful
CC      for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC      carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC      to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC      and other cancers in which SCCE is overexpressed. The present sequence is
CC      human SCCE peptide.
XX
SQ      Sequence 9 AA;
Query Match      90.0%; Score 36; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 LLPLQLL 8
Db      2 LLPLQLL 9
|||||:|||||
RESULT 10
ABB43858
ID      ABB43858 standard; Peptide; 23 AA.
XX
AC      ABB43858;
XX
DT      04-FEB-2002 (first entry)
XX
DE      Peptide #11364 encoded by human foetal liver single exon probe.
XX
KW      Human; foetal liver; gene expression; single exon nucleic acid probe.
XX
OS      Homo sapiens.

```

```

XX      WO200157277-A2.
PN
XX
PD
XX
PF      09-AUG-2001.
XX
XX      30-JAN-2001; 2001WO-US00669.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.
PR      21-SEP-2000; 2000US-0234687.
PR      27-SEP-2000; 2000US-0236359.
PR      04-OCT-2000; 2000GB-0024263.
XX
PA      (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI      Penn SG, Hanzel DK, Chen W, Rank DR;
XX
XX      WPI; 2001-483447/52.
XX
PT      Human genome-derived single exon nucleic acid probes useful for
PT      analyzing gene expression in human fetal liver -
XX
PS      Claim 27; SEQ ID NO 36493; 639pp + sequence listing; English.
XX
CC      The invention relates to a single exon nucleic acid probe for
CC      measuring human gene expression in a sample derived from human foetal
CC      liver. The single exon nucleic acid probes may be used for predicting,
CC      measuring and displaying gene expression in samples derived from human
CC      fetal liver. The present sequence is a peptide encoded by a single exon
CC      nucleic acid probe of the invention.
CC
CC      Note: The sequence data for this patent did not form part of the
CC      printed specification, but was obtained in electronic format directly
CC      from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ      Sequence 23 AA;
Query Match      87.5%; Score 35; DB 22; Length 23;
Best Local Similarity 77.8%; Pred. No. 6.2;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy      1 LLPLQLLL 9
Db      11 LLPLQLLLV 19
|||||:|||||
RESULT 11
AAM64837
ID      AAM64837 standard; Protein; 23 AA.
XX
AC      AAM64837;
XX
DT      05-NOV-2001 (first entry)
XX
DE      Human brain expressed single exon probe encoded protein SEQ ID NO: 36942.
XX
KW      Human; brain expressed exon; gene expression analysis; probe;
KW      microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW      epilepsy; cancer.
XX
OS      Homo sapiens.
XX
PN      WO200157275-A2.
XX
PD      09-AUG-2001.
XX
PF      30-JAN-2001; 2001WO-US00667.
XX
PR      04-FEB-2000; 2000US-0180312.
PR      26-MAY-2000; 2000US-0207456.
PR      30-JUN-2000; 2000US-0608408.
PR      03-AUG-2000; 2000US-0632366.

```

PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-483446/52.
 XX Single exon nucleic acid probes for analyzing gene expression in human
 PT brains -
 XX Example 4; SEQ ID NO: 36942; 650pp + Sequence Listing; English.
 PS The present invention provides a number of single exon nucleic acid
 CC probes which are derived from genomic sequences expressed in the human
 CC brain. They can be used to measure gene expression in brain cell samples,
 CC which may enable the diagnosis and improved treatment of nervous system
 CC diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
 CC epilepsy and cancers. The present sequence is a protein encoded by one of
 CC the probes of the invention.
 XX
 XX Sequence 23 AA;
 SQ
 Query Match 87.5%; Score 35; DB 22; Length 23;
 Best Local Similarity 77.8%; Pred. No. 6.2;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLPLQILLL 9
 Db 11 LLPLQLLLV 19
 RESULT 12
 AAM37771
 ID AAM37771 standard; Protein; 23 AA.
 XX AC
 XX AAM37771;
 XX 17-OCT-2001 (first entry)
 DT
 XX Peptide #11808 encoded by probe for measuring placental gene expression.
 DE
 XX Probe; microarray; human; placenta; antenatal diagnosis;
 KW genetic disorder.
 KW Homo sapiens.
 OS
 XX W0200157272-A2.
 PN
 XX 09-AUG-2001.
 PD
 XX 30-JAN-2001; 2001WO-US00663.
 XX
 PR 04-FEB-2000; 2000US-0180312.
 PR 26-MAY-2000; 2000US-0207456.
 PR 30-JUN-2000; 2000US-0608408.
 PR 03-AUG-2000; 2000US-0632366.
 PR 21-SEP-2000; 2000US-0234687.
 PR 27-SEP-2000; 2000US-0236359.
 PR 04-OCT-2000; 2000GB-0024263.
 XX (MOLE-) MOLECULAR DYNAMICS INC.
 XX Penn SG, Hanzel DK, Chen W, Rank DR;
 XX WPI; 2001-488897/53.
 XX Human genome-derived single exon nucleic acid probes useful for
 PT analyzing gene expression in human placenta -
 XX Claim 27; SEQ ID No 38040; 654pp; English.

XX The present invention relates to single exon nucleic acid probes (SENP;
 CC see AAI31315-AAI57546). The present sequence is a peptide encoded by one
 CC such probe. The probes are useful for producing a microarray for
 CC predicting, measuring and displaying gene expression in samples derived
 CC from human placenta. The probes are useful for antenatal diagnosis of
 CC human genetic disorders.
 XX
 XX Sequence 23 AA;
 SQ
 Query Match 87.5%; Score 35; DB 22; Length 23;
 Best Local Similarity 77.8%; Pred. No. 6.2;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 LLPLQILLL 9
 Db 11 LLPLQLLLV 19
 RESULT 13
 AAW98431
 ID AAW98431 standard; Protein; 480 AA.
 XX AC
 XX AAW98431;
 XX 31-MAR-1999 (first entry)
 DT
 XX H. pylori GHPO 446 protein.
 DE
 XX GHPO protein; Helicobacter infection; gastroduodenal disease; gastritis;
 KW peptic ulcer disease.
 KW Helicobacter pylori.
 OS
 XX W09843478-A1.
 PN
 XX 08-OCT-1998.
 PD
 XX 01-APR-1998; 98WO-US06371.
 XX
 PR 29-JUL-1997; 97US-0902615.
 PR 01-APR-1997; 97US-0833457.
 PR 24-JUN-1997; 97US-0881227.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (INMR) MERIEUX ORAVAX PASTEUR MERIEUX SERUMS.
 XX Al-Garawi A, Kleanthous H, Miller C, Oomen RP, Tomb J;
 PI WPI; 1998-542293/46.
 XX N-PSDB; AAX14150.
 DR
 XX New isolated Helicobacter polynucleotides - used to develop products
 PT for the diagnosis, prevention and treatment of Helicobacter
 PT infections and gastrointestinal diseases
 XX Claim 8; Page 730-733; 2054pp; English.
 PS This sequence represents a Helicobacter pylori GHPO protein of the
 CC invention. The polypeptides can be used for preventing or treating
 CC Helicobacter infections, and gastroduodenal diseases associated with
 CC these infections, including acute, chronic, and atrophic gastritis, and
 CC peptic ulcer diseases, e.g. gastric and duodenal ulcers. They can also be
 CC used for the production of antibodies. The products can also be used for
 CC detection and diagnosis.
 XX
 XX Sequence 480 AA;
 SQ
 Query Match 87.5%; Score 35; DB 19; Length 480;
 Best Local Similarity 88.9%; Pred. No. 1.4e+02;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 LLPLQILLL 9

Db 238 LLPLSILL 246
 RESULT 14
 AAY36291
 ID AAY36291 standard; Protein; 70 AA.
 XX
 AC AAY36291;
 XX
 DT 17-SEP-1999 (first entry)
 XX
 DE Human secreted protein encoded by gene 68.
 XX
 KW Human; secreted protein; cancer; tumour; developmental abnormality;
 KW foetal deficiency; blood disorder; immune system disorder; inflammation;
 KW autoimmune disease; allergy; Alzheimer's disease; cognitive disorder;
 KW schizophrenia; arthritis; asthma; psoriasis; sepsis; skin disorder;
 KW atherosclerosis; diabetes; cardiovascular disorder; kidney disorder;
 KW digestive disorder; endocrine disorder; infection; AIDS.
 XX
 OS Homo sapiens.
 XX
 PN WO9931117-A1.
 XX
 PD 24-JUN-1999.
 XX
 PF 17-DEC-1998; 98WO-US27059.
 XX
 PR 19-DEC-1997; 97US-0068369.
 PR 18-DEC-1997; 97US-0068006.
 PR 18-DEC-1997; 97US-0068007.
 PR 18-DEC-1997; 97US-0068008.
 PR 18-DEC-1997; 97US-0068053.
 PR 18-DEC-1997; 97US-0068054.
 PR 18-DEC-1997; 97US-0068057.
 PR 18-DEC-1997; 97US-0068064.
 PR 19-DEC-1997; 97US-0070923.
 PR 19-DEC-1997; 97US-0068169.
 PR 19-DEC-1997; 97US-0068365.
 PR 19-DEC-1997; 97US-0068367.
 PR 19-DEC-1997; 97US-0068368.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Carter KC, Duan RD, Feng P, Ferrie AM, Florence C;
 PI Florence K, Greene JM, Janat F, Kyaw H, Moore PA;
 PI Ni J, Rosen CA, Ruben SM, Shi Y, Soppet DR, Wei Y;
 PI Yu G;
 XX
 DR WPI: 1999-418749/35.
 DR N-PSDB; AAX97983.
 XX
 PT New isolated human genes encoding secreted polypeptides
 PS
 PS Claim 11; Page 382; 537pp; English.
 XX
 CC AAX97916 to AAX98029 represent 110 isolated human secreted protein
 CC genes. AAY36224 to AAY36727 represent the secreted proteins encoded by
 CC the 110 human genes. The genes and their corresponding secreted
 CC polypeptides are useful for preventing, treating or ameliorating medical
 CC conditions, e.g. by protein or gene therapy. Also pathological conditions
 CC can be diagnosed by determining the amount of the new polypeptides in a
 CC sample or by determining the presence of mutations in the new genes.
 CC Specific uses are described for each of the 110 genes, based on which
 CC tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, tumours, developmental
 CC abnormalities and foetal deficiencies, blood disorders, diseases of the
 CC immune system, autoimmune diseases, inflammation, allergies, Alzheimer's
 CC and cognitive disorders, schizophrenia, arthritis, asthma, psoriasis,
 CC sepsis, skin disorders, atherosclerosis, diabetes, cardiovascular
 CC disorders, kidney disorders, digestive/endocrine disorders, infections
 CC and AIDS. The polypeptides are also useful for identifying their binding

CC partners. The sequences given in AAX97907 to AAX97915 and AAY36223 are
 CC used in the exemplification of the present invention.
 XX
 SQ Sequence 70 AA;
 Query Match 85.0%; Score 34; DB 20; Length 70;
 Best Local Similarity 87.5%; Pred. No. 29;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LPLQLILL 9
 Db 5 LPLQLILL 12
 RESULT 15
 AAW89030
 ID AAW89030 standard; Protein; 73 AA.
 XX
 AC AAW89030;
 XX
 DT 01-MAR-1999 (first entry)
 XX
 DE Polypeptide fragment encoded by gene 165.
 XX
 KW Human; secreted protein; fusion protein; gene therapy; protein therapy;
 KW diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia;
 KW developmental abnormality; foetal deficiency; blood; allergy; renal;
 KW immune system; asthma; lymphocytic disease; brain; hepatitis; lymphoma;
 KW inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS;
 KW cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus;
 KW osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion;
 KW endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.
 XX
 OS Homo sapiens.
 XX
 PN WO9854963-A2.
 XX
 PD 10-DEC-1998.
 XX
 PF 04-JUN-1998; 98WO-US11422.
 XX
 PR 18-DEC-1997; 97US-0070923.
 PR 06-JUN-1997; 97US-0048877.
 PR 06-JUN-1997; 97US-0048881.
 PR 06-JUN-1997; 97US-0048884.
 PR 06-JUN-1997; 97US-0048893.
 PR 06-JUN-1997; 97US-0048896.
 PR 06-JUN-1997; 97US-0048899.
 PR 06-JUN-1997; 97US-0048915.
 PR 06-JUN-1997; 97US-0048949.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048972.
 PR 06-JUN-1997; 97US-0049020.
 PR 06-JUN-1997; 97US-0049375.
 PR 05-SEP-1997; 97US-0057628.
 PR 05-SEP-1997; 97US-0057635.
 PR 05-SEP-1997; 97US-0057644.
 PR 05-SEP-1997; 97US-0057647.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057661.
 PR 05-SEP-1997; 97US-0057667.
 PR 05-SEP-1997; 97US-0057761.
 PR 05-SEP-1997; 97US-0057764.
 PR 05-SEP-1997; 97US-0057770.
 PR 05-SEP-1997; 97US-0057775.
 PR 05-SEP-1997; 97US-0057778.
 PR 06-JUN-1997; 97US-0048875.
 PR 06-JUN-1997; 97US-0048878.
 PR 06-JUN-1997; 97US-0048882.
 PR 06-JUN-1997; 97US-0048885.
 PR 06-JUN-1997; 97US-0048894.
 PR 06-JUN-1997; 97US-0048897.
 PR 06-JUN-1997; 97US-0048900.

PR 06-JUN-1997; 97US-0048916.
 PR 06-JUN-1997; 97US-0048962.
 PR 06-JUN-1997; 97US-0048970.
 PR 06-JUN-1997; 97US-0048974.
 PR 06-JUN-1997; 97US-0049373.
 PR 05-SEP-1997; 97US-0037584.
 PR 05-SEP-1997; 97US-0037589.
 PR 05-SEP-1997; 97US-0057642.
 PR 05-SEP-1997; 97US-0057645.
 PR 05-SEP-1997; 97US-0057648.
 PR 05-SEP-1997; 97US-0057651.
 PR 05-SEP-1997; 97US-0057662.
 PR 05-SEP-1997; 97US-0057666.
 PR 05-SEP-1997; 97US-0057762.
 PR 05-SEP-1997; 97US-0057765.
 PR 05-SEP-1997; 97US-0057771.
 PR 05-SEP-1997; 97US-0057776.
 PR 06-JUN-1997; 97US-0048876.
 PR 06-JUN-1997; 97US-0048880.
 PR 06-JUN-1997; 97US-0048883.
 PR 06-JUN-1997; 97US-0048892.
 PR 06-JUN-1997; 97US-0048895.
 PR 06-JUN-1997; 97US-0048898.
 PR 06-JUN-1997; 97US-0048901.
 PR 06-JUN-1997; 97US-0048917.
 PR 06-JUN-1997; 97US-0048963.
 PR 06-JUN-1997; 97US-0048971.
 PR 06-JUN-1997; 97US-0049019.
 PR 06-JUN-1997; 97US-0049374.
 PR 05-SEP-1997; 97US-0057627.
 PR 05-SEP-1997; 97US-0057634.
 PR 05-SEP-1997; 97US-0057643.
 PR 05-SEP-1997; 97US-0057646.
 PR 05-SEP-1997; 97US-0057649.
 PR 05-SEP-1997; 97US-0057654.
 PR 05-SEP-1997; 97US-0057666.
 PR 05-SEP-1997; 97US-0057760.
 PR 05-SEP-1997; 97US-0057763.
 PR 05-SEP-1997; 97US-0057769.
 PR 05-SEP-1997; 97US-0057774.
 PR 05-SEP-1997; 97US-0057777.
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
 PI Fan P, Feng P, Ferrle AM, Fischer CL, Florence C;
 PI Florence K, Greene JW, Hu J, Kyaw H, Lafleur DW;
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
 XX WPI; 1999-059865/05.
 DR N-PSDB; AAV84575.
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Disclosure; Page 129; 772pp; English.
 XX
 CC The invention relates to nucleic acid sequences (AAV84411 to AAV84633)
 CC encoding human secreted proteins (AAW88534 to AAW88756). The secreted
 CC protein gene sequences are deposited with the ATCC under deposit numbers
 CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
 CC cells comprising recombinant vectors containing the nucleic acid
 CC sequences are used for the recombinant production of the secreted
 CC proteins. The polynucleotide and amino acid sequences are useful for are
 CC useful for preventing, treating or ameliorating medical conditions e.g.
 CC by protein or gene therapy. Pathological conditions can be also
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing

CC products for the diagnosis or treatment of cancer, neurodegenerative
 CC disorders, developmental abnormalities and foetal deficiencies, blood
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
 CC such as osteoporosis, arthritis or malignancies, diseases of testes,
 CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The present sequence represents a polypeptide fragment encoded by a
 CC gene of the invention (see descriptor line for gene number).
 XX
 SQ Sequence 73 AA;
 Query Match 85.0%; Score 34; DB 20; Length 73;
 Best Local Similarity 87.5%; Pred. No. 31;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LPLQLLLL 9
 Db 25 LPLQLLLL 32
 |||||
 |||||
 RESULT 16
 ABB51130
 ID ABB51130 standard; Protein; 73 AA.
 XX
 AC ABB51130;
 XX
 DT 07-FEB-2002 (first entry)
 XX
 DE Human secreted protein encoded by gene 165 SEQ ID NO:1083.
 XX
 KW Human; secreted protein; immunomodulatory; antisclerotic; anti-HIV;
 KW dermatological; immunosuppressive; antiinflammatory; immunostimulant;
 KW cytostatic; cardiac; vascular; anti-angiogenic; ophthalmological;
 KW neuroprotective; neurotropic; anticonvulsant; antialzheimer's; vulnary;
 KW antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; HIV infection; cancer;
 KW human immunodeficiency virus; hyperproliferative disorder; wound healing;
 KW Gaucher's disease; cardiovascular disease; scimitar syndrome; chemotaxis;
 KW Chaga's cardiomyopathy; coronary arteriosclerosis; angioecic disorder;
 KW corneal graft neovascularisation; diabetic retinopathy; regeneration;
 KW neurological disorder; Huntington's chorea; Alzheimer's disease;
 KW Parkinson's disease; infectious disease; chromosome 6.
 XX
 OS Homo sapiens.
 XX
 PN WO200162891-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 21-FEB-2001; 2001WO-US05614.
 XX
 PR 24-FEB-2000; 2000US-184836P.
 PR 29-MAR-2000; 2000US-193170P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ni J, Ebner R, Lafleur DW, Moore PA, Olsen HS, Rosen CA;
 PI Ruben SM, Soppet DR, Young PE, Shi Y, Florence KA, Wei Y;
 PI Florence C, Hu J, Li Y, Kyaw H, Fischer CL, Ferrle AM, Fan P;
 PI Feng P, Endress GA, Dillon PJ, Carter KC, Brewer LA, Yu G;
 PI Zeng Z, Greene JM;
 XX WPI; 2001-625724/72.
 DR
 XX Nucleic acids encoding 207 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating, e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 XX
 PS Disclosure; Page 360; 1533pp; English.

CC ABB50301 to ABB51287 and ABA83194 to ABA83441 represent human secreted
 CC proteins (I) and polynucleotide (II) sequences. (I) and (II) have various
 CC activities based on the tissues and cells the genes are expressed in.
 CC Example of these activities include: immunomodulatory; antisclerotic;
 CC dermatological; immunosuppressive; antiinflammatory; immunostimulant;
 CC anti-HIV; cytostatic; cardiant; anti-angiogenic; ophthalmological;
 CC neuroprotective; nootropic; anticonvulsant; antialzheimers; vascular;
 CC antiparkinsonian; antimicrobial; and vulnerary. (I) and (II) can be used
 CC in gene therapy and vaccine production. (I) and (II) can be used in the
 CC prevention, diagnosis and treatment of immune disorders (e.g. multiple
 CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angiogenic
 CC disorders (e.g. corneal graft neovascularisation and diabetic
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
 CC for promoting wound healing, regeneration and/or chemotaxis. ABA83185 to
 CC ABA83193 and ABB50300 represent sequences used in the exemplification of
 CC the present invention.

SQ Sequence 73 AA;

Query Match 85.0%; Score 34; DB 22; Length 73;
 Best Local Similarity 87.5%; Pred. No. 31;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILLL 9
 ||||:||||
 Db 25 LPLQLLLL 32

RESULT 17

AA080397
 ID AAO08397 standard; Protein; 86 AA.

XX AAO08397;

XX 06-NOV-2001 (first entry)

DE Human polypeptide SEQ ID NO 22289.

XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

XX Homo sapiens.

OS WO200164835-A2.

PN 07-SEP-2001.

XX 26-FEB-2001; 2001WO-US04927.

XX 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR N-PSDB; AAI88328.

XX Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -

XX Claim 20; SEQ ID NO 22289; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
 CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX SQ Sequence 86 AA;

Query Match 85.0%; Score 34; DB 22; Length 86;
 Best Local Similarity 87.5%; Pred. No. 36;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILLL 9
 ||||:||||
 Db 39 LPLQLLLL 46

RESULT 18

AA083417
 ID AAB93417 standard; Protein; 224 AA.

XX AAB93417;

XX 26-JUN-2001 (first entry)

DE Human protein sequence SEQ ID NO:12629.

XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

XX EP1074617-A2.

XX 07-FEB-2001.

XX 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.

PR 27-AUG-1999; 99JP-0300253.

PR 11-JAN-2000; 2000JP-0118776.

PR 02-MAY-2000; 2000JP-0183767.

PR 09-JUN-2000; 2000JP-0241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Salto K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -

XX Claim 8; SEQ ID 12629; 2537pp + CD ROM; English.

XX The present invention describes primer sets for synthesising 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a

CC polynucleotide which comprises a 3'-end sequence, where the
 CC oligonucleotide comprises at least 15 nucleotides and the combination of
 CC the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesising polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.

XX
 SQ Sequence 224 AA;

Query Match 85.0%; Score 34; DB 22; Length 224;
 Best Local Similarity 87.3%; Pred. No. 95;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILL 9
 DB 176 LPLQILL 183

RESULT 19
 AAW22303
 ID AAW22303 standard; Protein; 267 AA.

XX
 AC AAW22303;

XX 24-NOV-1997 (first entry)

XX Rat CRTI.

XX Rat; CRTI; expression; liver cancer; tissue; antibody; probe;
 KW detection; assay; diagnosis.

XX Rattus rattus.

XX WO9710333-A1.

XX 20-MAR-1997.

XX 17-SEP-1996; 96WO-JP02654.

XX 30-AUG-1996; 96JP-0229469.

XX 14-SEP-1995; 95JP-0236264.

XX 27-NOV-1995; 95JP-0331023.

XX 20-JUN-1996; 96JP-0179885.

XX (SUME) SUMITOMO ELECTRIC IND CO.

XX Kishimoto T, Kokura K, Kumagai Y, Makino Y, Tamura T;

XX WPI; 1997-202229/18.

XX N-PSDB; AAT7337.

XX Proteins with elevated expression in liver cancer, and related DNA -
 PT for production of antibodies useful in the diagnosis and monitoring
 PT of liver cancer

XX Claim 1; Pages 69-70; 119pp; Japanese.

XX The present sequence is rat CRTI, which has an elevated
 CC expression in liver cancer tissue. Anti-CRTI antibodies or CRTI
 CC cDNA can be used to detect or assay for CRTI or CRTI mRNA,
 CC indicating in which tissues CRTI expression is elevated, and
 CC therefore allowing liver cancer to be diagnosed and monitored.

XX Sequence 267 AA;

Query Match 85.0%; Score 34; DB 18; Length 267;

Best Local Similarity 88.9%; Pred. No. 1.1e+02;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LPLQILL 9
 DB 188 LPLQILL 196

RESULT 20

AAB87732

ID AAB87732 standard; Protein; 301 AA.

XX
 AC AAB87732;

XX 16-MAY-2001 (first entry)

XX Human T2R02 amino acid sequence SEQ ID NO:3.

XX Human; rat; mouse; T2R; taste receptor; G-protein coupled receptor;
 KW taste transduction G-protein coupled receptor; identification; tongue;
 KW taste sensory neuron; taste cell; taste modulator; food;
 KW taste signalling pathway.

XX Homo sapiens.

XX WO200118050-A2.

XX 15-MAR-2001.

XX 08-SEP-2000; 2000WO-US24821.

XX 10-SEP-1999; 99US-0393634.

XX 22-FEB-2000; 2000US-0510332.

XX (REGC) UNIV CALIFORNIA.

XX (USSH) US DEPT HEALTH & HUMAN SERVICES.

XX Zuker CS, Adler JE, Ryba N, Mueller K, Hoon M;

XX WPI; 2001-211396/21.

XX N-PSDB; AAF92503.

XX Nucleic acids encoding the T2R family of G-protein coupled taste
 PT receptors, useful for identifying taste modulators that can be used in
 PT food and pharmaceutical industries to customize taste, for e.g. to
 PT decrease the bitter taste of food -

XX Claim 19; Page 161; 249pp; English.

XX AAF92502 to AAF92572 represent nucleic acids which encode taste
 CC transduction G-protein coupled receptors designated T2R proteins.
 CC AAB87731 to AAB87824 represent T2R proteins, and AAB87825 to AAB87830
 CC represent T2R family consensus sequences from the present invention.
 CC The T2R proteins are taste modulators. The nucleic acids are useful as
 CC probes for the identification of taste cells, as the nucleic acids are
 CC specifically expressed in taste cells. They also serve as tools for the
 CC generation of taste topographic maps that elucidate the relationship
 CC between the taste cells of the tongue and taste sensory neurons leading
 CC to taste centres in the brain. The taste modulators are useful for
 CC pharmacological and genetic modulation of taste signalling pathways.
 CC Modulatory compounds comprising T2R proteins can therefore be used in
 CC food and pharmaceutical industries to customise taste, for e.g. to
 CC decrease the bitter taste of food or drugs.

XX Sequence 301 AA;

Query Match 85.0%; Score 34; DB 22; Length 301;

Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 LPLQILL 9

DB 42 LMPQILL 50

RESULT 21

AAR75642
ID AAR75642 standard; Protein; 371 AA.

XX AAR75642;

XX 10-FEB-1996 (first entry)

XX Bovine conglutinin.

XX Conglutinin; therapeutic; diagnostic.

XX Bos taurus.

XX Key Location/Qualifiers

FT Peptide 1..20

FT Protein /note= "signal peptide"

FT Region 21..371

FT Region /note= "mature bovine conglutinin"

FT Region /note= "carbohydrate-binding region"

FT Region /note= "carbohydrate-binding region"

XX WO9516697-A1.

XX 22-JUN-1995.

XX 14-DEC-1994; 94WO-US14656.

XX 16-DEC-1993; 93US-0168458.

XX (IMMU-) APPLIED IMMUNE SCI INC.

XX (UYBO-) UNIV BOSTON.

XX Lee YM, Leiby KR, Okarma TB, Sastry K;

XX WPI; 1995-231510/30.

XX N-PSDB; AAQ92554.

XX New nucleic acid encoding bovine conglutinin and its fragments - and related vectors, host cell, etc. useful e.g. in modulating conglutinin expression or for selective removal of cpds., carrying particular carbohydrate residues.

XX Claim 1; Fig 1A-1B; 29pp; English.

XX Conglutinin, including the complete sequence, the mature protein or the carbohydrate-binding fragments, are used to bind cells carrying the C1q receptor. Solid supports are used to remove compounds (especially proteins or peptides, e.g. immune complexes) having a carbohydrate component able to bind to the lectin region of conglutinin, optionally with recovery of bound compounds.

SQ Sequence 371 AA;

Query Match 85.0%; Score 34; DB 16; Length 371;

Best Local Similarity 77.8%; Pred. NO. 1.6e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPQLILL 9

|||||:||||

Db 3 LLPLSVLL 11

RESULT 22

AAW89026

ID AAW89026 standard; Protein; 424 AA.

XX AAW89026;

XX

01-MAR-1999 (first entry)

Polypeptide fragment encoded by gene 165.

Human; secreted protein; fusion protein; gene therapy; protein therapy; diagnosis; tissue; cancer; tumour; neurodegenerative disorder; leukaemia; developmental abnormality; foetal deficiency; blood; allergy; renal; immune system; asthma; lymphocytic disease; brain; hepatic; lymphoma; inflammation; ischaemic shock; Alzheimer's disease; restenosis; AIDS; cognitive disorder; schizophrenia; prostate; obesity; osteoclast; thymus; osteoporosis; arthritis; testis; lung; thyroiditis; thyroid; digestion; endocrine; metabolism; regulation; malabsorption; gastritis; neoplasm.

Homo sapiens.

WO9854963-A2.

10-DEC-1998.

04-JUN-1998; 98WO-US11422.

18-DEC-1997; 97US-0070923.

06-JUN-1997; 97US-0048877.

06-JUN-1997; 97US-0048881.

06-JUN-1997; 97US-0048884.

06-JUN-1997; 97US-0048893.

06-JUN-1997; 97US-0048896.

06-JUN-1997; 97US-0048899.

06-JUN-1997; 97US-0048915.

06-JUN-1997; 97US-0048949.

06-JUN-1997; 97US-0048964.

06-JUN-1997; 97US-0048972.

06-JUN-1997; 97US-0049020.

06-JUN-1997; 97US-0049375.

05-SEP-1997; 97US-0057628.

05-SEP-1997; 97US-0057635.

05-SEP-1997; 97US-0057644.

05-SEP-1997; 97US-0057647.

05-SEP-1997; 97US-0057650.

05-SEP-1997; 97US-0057661.

05-SEP-1997; 97US-0057761.

05-SEP-1997; 97US-0057764.

05-SEP-1997; 97US-0057770.

05-SEP-1997; 97US-0057775.

05-SEP-1997; 97US-0057778.

06-JUN-1997; 97US-0048875.

06-JUN-1997; 97US-0048878.

06-JUN-1997; 97US-0048882.

06-JUN-1997; 97US-0048885.

06-JUN-1997; 97US-0048894.

06-JUN-1997; 97US-0048897.

06-JUN-1997; 97US-0048900.

06-JUN-1997; 97US-0048916.

06-JUN-1997; 97US-0048962.

06-JUN-1997; 97US-0048970.

06-JUN-1997; 97US-0048974.

06-JUN-1997; 97US-0049373.

05-SEP-1997; 97US-0057584.

05-SEP-1997; 97US-0057629.

05-SEP-1997; 97US-0057642.

05-SEP-1997; 97US-0057645.

05-SEP-1997; 97US-0057648.

05-SEP-1997; 97US-0057651.

05-SEP-1997; 97US-0057662.

05-SEP-1997; 97US-0057668.

05-SEP-1997; 97US-0057762.

05-SEP-1997; 97US-0057765.

05-SEP-1997; 97US-0057771.

05-SEP-1997; 97US-0057776.

06-JUN-1997; 97US-0048876.

06-JUN-1997; 97US-0048880.

06-JUN-1997; 97US-0048883.

PR 06-JUN-1997; 97US-0048892.
 PR 06-JUN-1997; 97US-0048895.
 PR 06-JUN-1997; 97US-0048898.
 PR 06-JUN-1997; 97US-0048901.
 PR 06-JUN-1997; 97US-0048917.
 PR 06-JUN-1997; 97US-0048963.
 PR 06-JUN-1997; 97US-0048971.
 PR 06-JUN-1997; 97US-0049019.
 PR 06-JUN-1997; 97US-0049374.
 PR 05-SEP-1997; 97US-0057627.
 PR 05-SEP-1997; 97US-0057634.
 PR 05-SEP-1997; 97US-0057643.
 PR 05-SEP-1997; 97US-0057646.
 PR 05-SEP-1997; 97US-0057649.
 PR 05-SEP-1997; 97US-0057654.
 PR 05-SEP-1997; 97US-0057656.
 PR 05-SEP-1997; 97US-0057760.
 PR 05-SEP-1997; 97US-0057763.
 PR 05-SEP-1997; 97US-0057769.
 PR 05-SEP-1997; 97US-0057774.
 PR 05-SEP-1997; 97US-0057777.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Brewer LA, Carter KC, Dillon PJ, Ebner R, Endress GA;
 PI Fan P, Feng P, Ferrie AM, Fischer CL, Florence C;
 PI Florence K, Greene JM, Hu J, Kyaw H, Lafleur DW;
 PI Li Y, Moore PA, Ni J, Olsen HS, Rosen CA, Ruben SM;
 PI Shi Y, Soppet DR, Wei Y, Young P, Yu G, Zeng Z;
 XX
 DR WPI: 1999-059865/05.
 DR N-PSDB; AAV84575.
 DR
 XX
 PT New isolated human genes and the secreted polypeptides they encode -
 PT useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Disclosure; Page 129; 772pp; English.
 XX
 CC The invention relates to nucleic acid sequences (AAV84411 to AAV84633)
 CC encoding human secreted proteins (AAV88534 to AAV88756). The secreted
 CC protein gene sequences are deposited with the ATCC under deposit numbers
 CC ATCC 97979, 97974, 97975, 97976, 97977, 209007, 209008, 209009, 209010,
 CC 209011, 209080, 209081, 209082, 209083, 209084, 209085, 209511. Host
 CC cells comprising recombinant vectors containing the nucleic acid
 CC sequences are used for the recombinant production of the secreted
 CC proteins. The polynucleotide and amino acid sequences are useful for are
 CC by protein or gene therapy. Pathological conditions can be also
 CC diagnosed by determining the amount of the new polypeptides in a sample
 CC or by determining the presence of mutations in the new polynucleotides.
 CC Specific uses are described for each of the polynucleotides, based on
 CC which tissues they are most highly expressed in, and include developing
 CC products for the diagnosis or treatment of cancer, neurodegenerative
 CC disorders, developmental abnormalities and foetal deficiencies, blood
 CC disorders, tumours, leukemias, diseases of the immune system, autoimmune
 CC diseases, hepatic and renal disease, lymphomas, inflammation, allergies,
 CC ischemic shock, Alzheimer's and cognitive disorders, schizophrenia,
 CC restenosis, prostate diseases, obesity, disorders involving osteoclasts
 CC such as osteoporosis, arthritis or malignancies, diseases of testes,
 CC lung or thymus, digestive/endocrine disorders, infections and AIDS. The
 CC polypeptides are also useful for identifying their binding partners.
 CC The present sequence represents a polypeptide fragment encoded by a
 CC gene of the invention (see descriptor line for gene number).
 XX
 SQ Sequence 424 AA;
 Query Match 85.0%; Score 34; DB 20; Length 424;
 Best Local Similarity 87.5%; Pred. NO. 1.8e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 2 LPLQLLLL 9
 |||||

Db 376 LPLQLLLL 383
 RESULT 23
 ABB51126
 ID ABB51126 standard; Protein; 424 AA.
 XX
 AC ABB51126;
 XX
 DT 07-FEB-2002 (first entry)
 XX
 DE Human secreted protein encoded by gene 165 SEQ ID NO:1079.
 XX
 KW Human; secreted protein; immunomodulatory; antisclerotic; anti-HIV;
 KW dermatological; immunosuppressive; antiinflammatory; immunostimulant;
 KW cytotatic; cardiac; vascular; anti-angiogenic; ophthalmological;
 KW neuroprotective; nootropic; anticonvulsant; antialzheimers; vulnerary;
 KW antiparkinsonian; antimicrobial; gene therapy; vaccine; immune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; HIV infection; cancer;
 KW human immunodeficiency virus; hyperproliferative disorder; wound healing;
 KW Gaucher's disease; cardiovascular disease; Scimitar syndrome; chemotaxis;
 KW Chaga's cardiomyopathy; coronary arteriosclerosis; angioleptic disorder;
 KW corneal graft neovascularisation; diabetic retinopathy; regeneration;
 KW neurological disease; Huntington's chorea; Alzheimer's disease;
 KW Parkinson's disease; infectious disease; chromosome 6.
 XX
 OS Homo sapiens.
 XX
 XX
 PN WO200162891-A2.
 XX
 PD 30-AUG-2001.
 XX
 PF 21-FEB-2001; 2001WO-US05614.
 XX
 PR 24-FEB-2000; 2000US-184836P.
 PR 29-MAR-2000; 2000US-193170P.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Ni J, Ebner R, Lafleur DW, Moore PA, Olsen HS, Rosen CA;
 PI Ruben SM, Soppet DR, Young PE, Shi Y, Florence KA, Wei Y;
 PI Florence C, Hu J, Li Y, Kyaw H, Fischer CL, Ferrie AM, Fan P;
 PI Feng P, Endress GA, Dillon PJ, Carter KC, Brewer LA, Yu G;
 PI Zeng Z, Greene JM;
 XX
 DR WPI: 2001-625724/72.
 XX
 XX Nucleic acids encoding 207 human secreted polypeptides, useful for
 PT preventing, diagnosing and/or treating, e.g. cancers, Parkinson's
 PT disease and diabetic retinopathy -
 PT
 PS Disclosure; Page 360; 1533pp; English.
 XX
 CC ABB50301 to ABB51287 and ABA83194 to ABA83441 represent human secreted
 CC proteins (I) and polynucleotide (II) sequences. (I) and (II) have various
 CC activities based on the tissues and cells the genes are expressed in.
 CC Example of these activities include: immunomodulatory; antisclerotic;
 CC dermatological; immunosuppressive; antinflammatory; immunostimulant;
 CC anti-HIV; cytostatic; cardiac; anti-angiogenic; ophthalmological;
 CC neuroprotective; nootropic; anticonvulsant; antialzheimers; vascular;
 CC antiparkinsonian; antimicrobial; and vulnerary. (I) and (II) can be used
 CC in gene therapy and vaccine production. (I) and (II) can be used in the
 CC prevention, diagnosis and treatment of immune disorders (e.g. multiple
 CC sclerosis, systemic lupus erythematosus and human immunodeficiency virus
 CC (HIV) infections), hyperproliferative disorders (e.g. cancers and
 CC Gaucher's disease), cardiovascular diseases (e.g. Scimitar syndrome,
 CC Chaga's cardiomyopathy and coronary arteriosclerosis), angioleptic
 CC disorders (e.g. corneal graft neovascularisation and diabetic
 CC retinopathy), neurological disorders (e.g. Huntington's chorea,
 CC Alzheimer's disease and Parkinson's disease), infectious diseases and/or
 CC for promoting wound healing, regeneration and/or chemotaxis. ABA83185 to
 CC ABA83193 and ABB50300 represent sequences used in the exemplification of
 CC the present invention.

```
XX SQ Sequence 424 AA;
Query Match 85.0%; Score 34; DB 22; Length 424;
Best Local Similarity 87.5%; Pred. No. 1.8e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQIULL 9
    IIII:III
Db 376 LPLQLLLLL 383

RESULT 24
AAM00784
ID AAM00784 standard; Protein; 1001 AA.
XX AC AAM00784;
XX DT 01-OCT-2001 (first entry)
XX DE Human bone marrow protein, SEQ ID NO: 147.
XX KW Human; bone marrow; antinflammatory; cytostatic; neuroprotective;
KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
KW immunosuppressive; gene therapy; cytokine cell proliferation;
KW cell differentiation modulator; immune disorder; infection; cancer;
KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.
XX OS Homo sapiens.
XX PN WO200153453-A2.
XX PD 26-JUL-2001.
XX PF 23-DEC-2000; 2000WO-US34960.
XX PR 21-JAN-2000; 2000US-0488725.
XX PR 25-APR-2000; 2000US-0552317.
XX PR 09-JUL-2000; 2000US-0598042.
XX PR 19-JUL-2000; 2000US-0620312.
XX PR 03-AUG-2000; 2000US-0653450.
XX PR 14-SEP-2000; 2000US-0662191.
XX PR 19-OCT-2000; 2000US-0693036.
XX PR 30-NOV-2000; 2000US-0250583.
XX PA (HYSE-) HYSEQ INC.
XX PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Drmanac RT;
XX WPI; 2001-488707/53.
XX DR N-PSDB; AAH89903.
XX PT Novel bone-marrow-expressed polynucleotides and polypeptides, useful
XX for treating e.g. cancer and immune deficiency disorders.
XX PS Claim 10; Page 311-313; 648pp; English.
XX CC The present sequence is one of 251 novel human polypeptides encoded
XX by a bone marrow-expressed polynucleotide. The polynucleotide and the
XX polypeptide encoded by it are useful in the treatment of various
XX immune deficiencies and disorders. The deficiencies and disorders may
XX be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
XX infection, or may result from an autoimmune disorder, a coagulation
XX disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
XX suppression of an inflammatory response or treatment of a nervous
XX system disorder such as Alzheimer's disease. Detection of the presence
XX or increased expression of the polynucleotide or the protein it
XX encodes is useful for the diagnosis and/or prognosis of one
XX or more types of cancer. The polynucleotide and polypeptide can be
XX used as nutritional sources or supplements and in the screening of
XX chemical compounds as potential drugs.

XX SQ Sequence 1001 AA;
Query Match 85.0%; Score 34; DB 22; Length 1001;
Best Local Similarity 87.5%; Pred. No. 4.4e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQIULL 9
    IIII:III
Db 951 LPLQLLLLL 958

RESULT 25
ABG09527
ID ABG09527 standard; Protein; 1080 AA.
XX AC ABG09527;
XX DT 13-FEB-2002 (first entry)
XX DE Novel human diagnostic protein #9518.
XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-639362/73.
XX DR N-PSDB; AAS73714.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits and to assess
XX biodiversity.
XX PS Claim 20; SEQ ID No 39886; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX and gene mapping, and in recombinant production of (II). The
XX polynucleotides are also used in diagnostics as expressed sequence tags
XX for identifying expressed genes. (I) is useful in gene therapy techniques
XX to restore normal activity of (II) or to treat disease states involving
XX (II). (II) is useful for generating antibodies against it, detecting or
XX quantitating a polypeptide in tissue, as molecular weight markers and as
XX a food supplement. (II) and its binding partners are useful in medical
XX imaging of sites expressing (II). (I) and (II) are useful for treating
XX disorders involving aberrant protein expression or biological activity.
XX The polypeptide and polynucleotide sequences have applications in
XX diagnostics, forensics, gene mapping, identification of mutations
XX responsible for genetic disorders or other traits to assess biodiversity
XX and to produce other types of data and products dependent on DNA and
XX amino acid sequences. ABG00010-ABG30377 represent novel human
XX diagnostic amino acid sequences of the invention.
XX CC Note: the sequence data for this patent did not appear in the printed
XX specification, but was obtained in electronic format directly from WIPO
XX at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 1080 AA;
```

Query Match 85.0%; Score 34; DB 22; Length 1080;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILLL 9
||||:||||
DB 1032 LPLQILLL 1039

RESULT 26
AAM00897
ID AAM00897 standard; Protein: 1080 AA.

XX AC AAM00897;

XX DT 01-OCT-2001 (first entry)

XX DE Human bone marrow protein, SEQ ID NO: 373.

XX KW Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
KW immunosuppressive; gene therapy; cytokine cell proliferation;
KW cell differentiation modulator; immune disorder; infection; cancer;
KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.

XX OS Homo sapiens.

XX PN WO200153453-A2.

XX PD 26-JUL-2001.

XX PF 23-DEC-2000; 2000WO-US34960.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PR 09-JUL-2000; 2000US-0598042.

XX PR 19-JUL-2000; 2000US-0620312.

XX PR 03-AUG-2000; 2000US-0653450.

XX PR 14-SEP-2000; 2000US-0662191.

XX PR 19-OCT-2000; 2000US-0693036.

XX PR 30-NOV-2000; 2000US-0250583.

XX PA (HYSE-) HYSEQ INC.

XX PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;

XX PI Ren F, Wang J, Werhman T, Xu C, Xue AJ, Yang Y, Zhang J;

XX PI Zhao QA, Zhou P, Drmanac RT;

XX DR WPI; 2001-488707/53.

XX DR N-PSDB; AAH90016.

XX PT Novel bone-marrow-expressed polynucleotides and polypeptides, useful
PT for treating e.g. cancer and immune deficiency disorders -

XX PS Claim 10; Page 459-461; 648pp; English.

XX CC The present sequence is one of 251 novel human polypeptides encoded
CC by a bone marrow-expressed polynucleotide. The polynucleotide and the
CC polypeptide encoded by it are useful in the treatment of various
CC immune deficiencies and disorders. The deficiencies and disorders may
CC be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal
CC infection, or may result from an autoimmune disorder, a coagulation
CC disorder (e.g. haemophilia), inhibition of tumour cell proliferation,
CC suppression of an inflammatory response or treatment of a nervous
CC system disorder such as Alzheimer's disease. Detection of the presence
CC or increased expression of the polynucleotide or the protein it
CC encodes is useful for the diagnosis and/or prognosis of one
CC or more types of cancer. The polynucleotide and polypeptide can be
CC used as nutritional sources or supplements and in the screening of
CC chemical compounds as potential drugs.

XX SQ Sequence 1080 AA;

Query Match 85.0%; Score 34; DB 22; Length 1080;
Best Local Similarity 87.5%; Pred. No. 4.7e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQILLL 9
||||:||||
DB 1032 LPLQILLL 1039

RESULT 27
AAB42131
ID AAB42131 standard; Protein: 1092 AA.

XX AC AAB42131;

XX DT 08-FEB-2001 (first entry)

XX DE Human ORFX ORF1895 polypeptide sequence SEQ ID NO:3790.

XX KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; neurotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; erythroloidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.

XX OS Homo sapiens.

XX PN WO200058473-A2.

XX PD 05-OCT-2000.

XX PF 31-MAR-2000; 2000WO-US08621.

XX PR 31-MAR-1999; 99US-0127607.

XX PR 02-APR-1999; 99US-0127636.

XX PR 05-APR-1999; 99US-0127728.

XX PA (CURA-) CURAGEN CORP.

XX PI Shimkets RA, Leach M;

XX DR WPI; 2000-602362/57.

XX DR N-PSDB; AAC76340.

XX PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -

XX PS Claim 11; Page 2940-2942; 5507pp; English.

XX CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; neurotropic; immunosuppressant;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; immunosuppressive;
CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy

CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX
 SQ Sequence 1092 AA;

Query Match 85.0%; Score 34; DB 21; Length 1092;
 Best Local Similarity 87.5%; Pred. No. 4.8e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLILL 9
 ||||:||||
 Db 1044 LPLQLLEL 1051

RESULT 28
 AAM39421
 ID AAM39421 standard; Protein: 1092 AA.
 XX
 AC AAM39421;
 XX
 DT 22-OCT-2001 (first entry)
 XX
 DE Human polypeptide SEQ ID NO 2566.
 XX
 KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
 KW peripheral nervous system; neuropathy; central nervous system; CNS;
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
 KW ankyrotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
 KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
 KW leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153312-A1.
 XX
 PD 26-JUL-2001.
 XX
 PF 26-DEC-2000; 2000WO-US34263.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 29-NOV-2000; 2000US-0727344.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
 DR WPI: 2001-442253/47.
 DR N-PSDB: AAI58577.
 XX
 XX Novel nucleic acids and polypeptides, useful for treating disorders
 PT such as central nervous system injuries -
 PS Example 4; SEQ ID NO 2566; 10078pp; English.
 XX
 XX The invention relates to human nucleic acids (AAI57798-AAI61369) and
 CC the encoded polypeptides (AAM38642-AAAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide

CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX
 SQ Sequence 1092 AA;

Query Match 85.0%; Score 34; DB 22; Length 1092;
 Best Local Similarity 87.5%; Pred. No. 4.8e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLILL 9
 ||||:||||
 Db 1044 LPLQLLEL 1051

RESULT 29
 AAM00996
 ID AAM00996 standard; Protein: 1139 AA.
 XX
 AC AAM00996;
 XX
 DT 01-OCT-2001 (first entry)
 XX
 DE Human bone marrow protein, SEQ ID NO: 497.
 XX
 KW Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
 KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
 KW immunosuppressive; gene therapy; cytokine cell proliferation;
 KW cell differentiation modulator; immune disorder; infection; cancer;
 KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.
 XX
 OS Homo sapiens.
 XX
 PN WO200153453-A2.
 XX
 PD 26-JUL-2001.
 XX
 PF 23-DEC-2000; 2000WO-US34960.
 XX
 PR 21-JAN-2000; 2000US-0488725.
 PR 25-APR-2000; 2000US-0552317.
 PR 09-JUL-2000; 2000US-0598042.
 PR 19-JUL-2000; 2000US-0620312.
 PR 03-AUG-2000; 2000US-0653450.
 PR 14-SEP-2000; 2000US-0662191.
 PR 19-OCT-2000; 2000US-0693036.
 PR 30-NOV-2000; 2000US-0250583.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
 PI Ren F, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
 PI Zhao QA, Zhou P, Drmanac RT;
 DR WPI: 2001-488707/53.
 DR N-PSDB: AAH90115.
 XX
 XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful
 PT for treating e.g. cancer and immune deficiency disorders -
 PS Claim 10; Page 641-643; 648pp; English.
 XX
 XX The present sequence is one of 251 novel human polypeptides encoded
 CC by a bone marrow-expressed polynucleotide. The polynucleotide and the

polypeptide encoded by it are useful in the treatment of various immune deficiencies and disorders. The deficiencies and disorders may be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal infection, or may result from an autoimmune disorder, a coagulation disorder (e.g. haemophilia), inhibition of tumour cell proliferation, suppression of an inflammatory response or treatment of a nervous system disorder such as Alzheimer's disease. Detection of the presence or increased expression of the polynucleotide or the protein it encodes is useful for the diagnosis and/or prognosis of one or more types of cancer. The polynucleotide and polypeptide can be used as nutritional sources or supplements and in the screening of chemical compounds as potential drugs.

XX
SQ Sequence 1139 AA;

Query Match 85.0%; Score 34; DB 22; Length 1139;
Best Local Similarity 87.5%; Pred. No. 5e+02; Mismatches 0; Gaps 0;
Matches 7; Conservative 1;

OY 2 LPLQLLLL 9
||||:||||
Db 1091 LPLQLLLL 1098

RESULT 30
AAM00997
ID AAM00997 standard; Protein; 1153 AA.
XX
AC AAM00997;
XX
DT 01-OCT-2001 (first entry)
XX
DE Human bone marrow protein, SEQ ID NO: 498.
XX
KW Human; bone marrow; antiinflammatory; cytostatic; neuroprotective;
KW antiviral; antibacterial; antifungal; anti-HIV; haemostatic;
KW immunosuppressive; gene therapy; cytokine cell proliferation;
KW cell differentiation modulator; immune disorder; infection; cancer;
KW human immunodeficiency virus; HIV; autoimmune disorder; haemophilia.
XX
OS Homo sapiens.
XX
PN WO200153453-A2.
XX
PD 26-JUL-2001.
XX
PF 23-DEC-2000; 2000WO-US34960.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 30-NOV-2000; 2000US-0250583.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Ford JE, Boyle BJ, Tang YT, Liu C, Asundi V, Chen R, Ma Y;
PI Ren F, Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Drmanac RT;
XX
DR WPI: 2001-488707/53.
DR N-PSDB; AAH90116.
XX
XX Novel bone-marrow-expressed polynucleotides and polypeptides, useful
PT for treating e.g. cancer and immune deficiency disorders -
XX
XX Claim 10; Page 643-646; 648pp; English.
XX
XX The present sequence is one of 251 novel human polypeptides encoded
CC by a bone marrow-expressed polynucleotide. The polynucleotide and the

polypeptide encoded by it are useful in the treatment of various immune deficiencies and disorders. The deficiencies and disorders may be genetic, may be caused by a viral (e.g. HIV), bacterial or fungal infection, or may result from an autoimmune disorder, a coagulation disorder (e.g. haemophilia), inhibition of tumour cell proliferation, suppression of an inflammatory response or treatment of a nervous system disorder such as Alzheimer's disease. Detection of the presence or increased expression of the polynucleotide or the protein it encodes is useful for the diagnosis and/or prognosis of one or more types of cancer. The polynucleotide and polypeptide can be used as nutritional sources or supplements and in the screening of chemical compounds as potential drugs.

XX
SQ Sequence 1153 AA;

Query Match 85.0%; Score 34; DB 22; Length 1153;
Best Local Similarity 87.5%; Pred. No. 5e+02; Mismatches 0; Gaps 0;
Matches 7; Conservative 1;

OY 2 LPLQLLLL 9
||||:||||
Db 1105 LPLQLLLL 1112

RESULT 31
AAM41207
ID AAM41207 standard; Protein; 1314 AA.
XX
AC AAM41207;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 6138.
XX
KW Human; neutropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB; AA160363.
XX
XX Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Example 2; SEQ ID NO 6138; 10078pp; English.
XX
XX The invention relates to human nucleic acids (AA157798-AA161369) and

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
 CC immunosuppressant and cytostatic activity. The polynucleotides are useful
 CC in gene therapy. A composition containing a polypeptide or polynucleotide
 CC of the invention may be used to treat diseases of the peripheral nervous
 CC system, such as peripheral nervous injuries, peripheral neuropathy and
 CC localised neuropathies and central nervous system diseases, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
 CC utilisation of the activities such as: Immune system suppression,
 CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
 CC assays for receptor activity, arthritis and inflammation, leukaemias and
 CC C.N.S disorders.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification.

XX Sequence 1214 AA;
 SQ Query Match 85.0%; Score 34; DB 22; Length 1214;
 Best Local Similarity 87.5%; Pred. NO. 5.3e+02;
 Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLILL 9
 ||||:||||
 Db 1166 LPLQLILL 1173

RESULT 32
 AAU47328
 ID AAU47328 standard; Protein: 60 AA.

XX AC AAU47328;
 XX DT 27-FEB-2002 (first entry)

DE DE Propionibacterium acnes immunogenic protein #8224.

XX KW SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
 KW uveitis; endophthalmitis; bone; joint; central nervous system; ELISA;
 KW inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
 KW dermatological; osteopathic; neuroprotectant.

XX OS Propionibacterium acnes.

PN WO200181581-A2.

XX PD 01-NOV-2001.

XX PF 20-APR-2001; 2001WO-US12865.

XX PR 21-APR-2000; 2000US-199047P.

XX PR 02-JUN-2000; 2000US-208841P.

XX PR 07-JUL-2000; 2000US-216747P.

XX PA (CORI-) CORIXA CORP.

XX PI Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;

XX PI L'maisonneuve J, Zhang Y, Jen S, Carter D;

XX DR WPI; 2001-616774/71.

XX DR N-PSDB; AAS59538.

XX PT Propionibacterium acnes polypeptides and nucleic acids useful for
 PT vaccinating against and diagnosing infections, especially useful for
 PT treating acne vulgaris -

XX PS Example 1; SEQ ID No 8523; 1069pp; English.

XX CC Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
 CC polypeptides. The proteins and their associated DNA sequences are used in
 CC the treatment, prevention and diagnosis of medical conditions caused by
 CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
 CC pustulosis, hypertosis and osteomyelitis), uveitis and endophthalmitis.

CC P. acnes is also involved in infections of bone, joints and the central
 CC nervous system, however it is particularly involved in the inflammatory
 CC lesions associated with acne vulgaris. A method for detecting the
 CC presence or absence of P. acnes in a patient comprises contacting a
 CC sample with a binding agent that binds to the proteins of the invention
 CC and determining the amount of bound protein in the sample. The
 CC polypeptides may be used as antigens in the production of antibodies
 CC specific for P. acnes proteins. These antibodies can be used to
 CC downregulate expression and activity of P. acnes polypeptides and
 CC therefore treat P. acnes infections. The antibodies may also be used as
 CC diagnostic agents for determining P. acnes presence, for example, by
 CC enzyme linked immunosorbent assay (ELISA).

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 60 AA;

XX Query Match 82.5%; Score 33; DB 22; Length 60;
 Best Local Similarity 75.0%; Pred. NO. 38;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LPLQLILL 9
 ||||:||||
 Db 19 LPLQVILL 26

RESULT 33
 AA012472

ID AA012472 standard; Protein: 61 AA.

XX AC AA012472;

XX DT 06-NOV-2001 (first entry)

DE DE Human polypeptide SEQ ID NO 26364.

XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation.

XX OS Homo sapiens.

PN WO200164835-A2.

XX PD 07-SEP-2001.

XX PF 26-FEB-2001; 2001WO-US04927.

XX PR 28-FEB-2000; 2000US-0515126.

XX PR 18-MAY-2000; 2000US-0577409.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX XX WPI; 2001-514838/56.

XX DR N-PSDB; AA192403.

XX PT Isolated nucleic acids and polypeptides, useful for preventing
 PT diagnosing and treating e.g. leukaemia, inflammation and immune
 PT disorders -

XX PS Claim 20; SEQ ID NO 26364; 1399pp + Sequence Listing; English.

XX CC The invention relates to human polynucleotides (AA179941-AA193841) and
 CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 61 AA;

Query Match 82.5%; Score 33; DB 22; Length 61;
Best Local Similarity 87.5%; Pred. No. 39;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 8
Db 5 LLPLQILLV 12
|||||:

RESULT 34
AAB23633
ID AAB23633 standard; Protein; 77 AA.
XX
AC AAB23633;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human secreted protein SEQ ID NO: 89.
XX
KW Human secreted protein; cytokine; cell proliferation;
KW nutritional supplement; immune modulation; autoimmune disorder;
KW haematopoiesis regulation; tissue growth; haemostasis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200049134-A1.
XX
PD 24-AUG-2000.
XX
PF 18-FEB-2000; 2000WO-US04340.
XX
PR 19-FEB-1999; 99US-0120680.
PR 23-APR-1999; 99US-0298733.
PR 17-AUG-1999; 99US-0149639.
PR 23-SEP-1999; 99US-0155686.
PR 01-OCT-1999; 99US-0157247.
PR 29-NOV-1999; 99US-0167822.
PR 29-NOV-1999; 99US-0167823.
PR 15-FEB-2000; 2000US-0298733.
XX
PA (ALPH-) ALPHAGENE INC.
XX
PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;
XX
DR WPI; 2000-549267/50.
XX
PT New secreted proteins and polynucleotides encoding them, which are
PT derived from Homo sapiens, useful for therapy, diagnosis, and research,
PT as well as nutritional sources or supplements -
XX
PS Disclosure; Page 295-296; 309pp; English.
XX
CC The present invention is concerned with a number of secreted proteins
CC and their coding sequences isolated from various human cDNA libraries.
CC The proteins and coding sequences can be used in the isolation of
CC similar genes and proteins, in the elucidation of their function in vivo,
CC and to treat a number of conditions. It is possible that they may have
CC uses as nutritional supplements, as cytokine or cell proliferation
CC factors, in immune modulation, where they may be used to treat immune and
CC autoimmune diseases, as haematopoiesis regulators (treating myeloid or
CC lymphoid cell deficiencies), in the promotion of tissue growth, they may
CC have chemokine or chemotactic activity, haemostatic or thrombolytic
CC activity, or anti-inflammatory activity. No information about sequences

CC AAB23632-B23645 is given in the specification.
XX
SQ Sequence 77 AA;

Query Match 82.5%; Score 33; DB 21; Length 77;
Best Local Similarity 77.8%; Pred. No. 49;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
Db 7 LLPLHLLL 15
||||:||||

RESULT 35
AAU14538
ID AAU14538 standard; Protein; 146 AA.
XX
AC AAU14538;
XX
DT 24-OCT-2001 (first entry)
XX
DE Human novel protein #409.
XX
KW Human; novel protein; Antianaemic; osteopathic; antiinflammatory;
KW immunomodulatory; cytostatic; neuroprotective; vulnary; nootropic;
KW anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;
KW antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;
KW thrombolytic; immunogen; antibody; gene therapy; neurological disorder;
KW Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;
KW tissue regeneration; Immune disorder.
XX
OS Homo sapiens.
XX
PN WO200155437-A2.
XX
PD 02-AUG-2001.
XX
PF 25-JAN-2001; 2001WO-US02623.
XX
PR 25-JAN-2000; 2000US-0491404.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-451939/48.
DR N-PSDB; AAS22843.
XX
PT Isolated polypeptides useful for treating anti-inflammatory diseases,
PT nervous system disorders, and for regenerating bone and cartilage -
XX
PS Example 4; Page 864; 894pp; English.
XX
CC The invention relates to polynucleotides encoding novel human
CC proteins or their active domains. The polypeptides, polynucleotides and
CC antibodies raised against the polypeptides are used in a method of
CC treatment of a mammal and prevention of disorders caused by the aberrant
CC protein expression or activity. The polypeptides can be used as
CC molecular weight markers, food supplements, and in antibody production.
CC The polypeptides are used to identify compounds which bind to the
CC polypeptides. Polynucleotides of the invention are used as probes and
CC primers, for sequencing, for chromosome or gene mapping, in the
CC production of recombinant proteins, and in generating anti-sense DNA or
CC RNA and in gene therapy. Polypeptides of the invention can be used to
CC target drugs to a tumour, in assays to determine biological activity, to
CC raise antibodies/elicit an immune response, to determine quantitative
CC protein levels, as tissue markers, and to isolate receptors or ligands.
CC Polypeptides of the invention may also be useful in treating platelet
CC disorders, stem cell disorders, regenerating bone, cartilage, tendon,
CC ligament and/or nerve tissue, wound healing, treating burns, promoting
CC the proliferation, differentiation and survival of stem cells, as a
CC contraceptive, treating osteoporosis and osteoarthritis, anaemia,
CC Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral

CC sclerosis, stroke, immune deficiencies resulting from bacterial, viral or
CC fungal infection or from autoimmunity, cancer, allergy, asthma,
CC graft-versus-host disease, eczema, haemophilia, thrombosis,
CC anti-inflammatory diseases, nervous system disorders, and infection.
XX The present sequence represents a protein of the invention.
SQ Sequence 146 AA;

Query Match 82.5%; Score 33; DB 22; Length 146;
Best Local Similarity 77.8%; Pred. No. 94;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9

||||:||||

Db 7 LLPLHLLL 15

RESULT 36

ABB69455

ID ABB69455 standard; Protein: 149 AA.

XX ABB69455;

AC ABB69455;

XX 26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 35157.

XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.

XX Drosophila melanogaster.

XX WO200171042-A2.

XX 27-SEP-2001.

XX 23-MAR-2001; 2001WO-US09231.

XX 23-MAR-2000; 2000US-191637P.

PR 11-JUL-2000; 2000US-0614150.

XX (PEKE) PE CORP NY.

PI Venter JC, Adams M, Li PWD, Myers EW;

XX WPI; 2001-656860/75.

DR N-PSDB; ABL13558.

XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -

XX Disclosure; SEQ ID NO 35157; 2lpp + Sequence Listing; English.

XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 149 AA;

Query Match

Best Local Similarity 82.5%; Score 33; DB 22; Length 149;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9

Db 41 MLPELLLL 49

:|||::|||

RESULT 37

AAU14302

ID AAU14302 standard; Protein: 181 AA.

XX AAU14302;

AC AAU14302;

XX 24-OCT-2001 (first entry)

DT Human novel protein #173.

XX Human; novel protein; Antianaemic; osteopathic; antiinflammatory;

XX immunomodulatory; cytostatic; neuroprotective; vulnerary; nootropic;

XX anticonvulsant; antiarthritic; cerebroprotective; antifungal; antiviral;

XX antibacterial; antiallergic; dermatological; haemostatic; antiasthmatic;

XX thrombolytic; immunogen; antibody; gene therapy; neurological disorder;

XX Parkinson's disease; inflammatory disorder; cancer; asthma; osteoporosis;

XX tissue regeneration; immune disorder.

XX Homo sapiens.

OS WO200155437-A2.

XX 02-AUG-2001.

PD 25-JAN-2001; 2001WO-US02623.

XX 25-JAN-2000; 2000US-0491404.

XX (HYSE-) HYSEQ INC.

PA Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-451939/48.

XX N-PSDB; AAS22607.

XX Isolated polypeptides useful for treating anti-inflammatory diseases,

PT nervous system disorders, and for regenerating bone and cartilage -

XX Example 4; Page 622; 894pp; English.

PS The invention relates to polynucleotides encoding novel human

XX proteins or their active domains. The polypeptides, polynucleotides and

XX antibodies raised against the polypeptides are used in a method of

XX treatment of a mammal and prevention of disorders caused by the aberrant

XX protein expression or activity. The polypeptides can be used as

XX molecular weight markers, food supplements, and in antibody production.

XX The polypeptides are used to identify compounds which bind to the

XX polypeptides. Polynucleotides of the invention are used as probes and

XX primers, for sequencing, for chromosome or gene mapping, in the

XX production of recombinant proteins, and in generating anti-sense DNA or

XX RNA and in gene therapy. Polypeptides of the invention can be used to

XX target drugs to a tumour, in assays to determine biological activity, to

XX raise antibodies/elicit an immune response, to determine quantitative

XX protein levels, as tissue markers, and to isolate receptors or ligands.

XX Polypeptides of the invention may also be useful in treating platelet

XX disorders, stem cell disorders, regenerating bone, cartilage, tendon,

XX ligament and/or nerve tissue, wound healing, treating burns, promoting

XX the proliferation, differentiation and survival of stem cells, as a

XX contraceptive, treating osteoporosis and osteoarthritis, anaemia, as

XX Alzheimer's, Parkinson's and Huntington's diseases, amyotrophic lateral

XX sclerosis, stroke, immune deficiencies resulting from bacterial, viral or

XX fungal infection or from autoimmunity, cancer, allergy, asthma,

XX graft-versus-host disease, eczema, haemophilia, thrombosis,

XX anti-inflammatory diseases, nervous system disorders, and infection.

XX The present sequence represents a protein of the invention.

XX Sequence 181 AA;

SQ Query Match 82.5%; Score 33; DB 22; Length 181;

```

Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
Db 7 LLPLHLLL 15

RESULT 38
AAR98922
ID AAR98922 standard; Protein: 190 AA.
XX
AC AAR98922;
XX
DT 29-OCT-1996 (first entry)
XX
DE Murine APLP1 (aa21-210).
XX
KW APLP1; amyloid precursor-like protein 1; APP;
KW amyloid beta-protein precursor; promoter; transcription;
KW upstream stimulatory factor; Alzheimer's disease.
XX
OS Mus sp.
XX
PN W09615265-A1.
XX
PD 23-MAY-1996.
XX
PF 09-NOV-1995; 95WO-US14416.
XX
PR 10-NOV-1994; 94US-0339152.
XX
PA (GEO ) GEN HOSPITAL CORP.
XX
PI Kovacs DM, Tanzi RE;
XX
DR WPI; 1996-259860/26.
XX
PT Modulating transcription from amyloid beta-protein precursor
PT promoter - using upstream stimulatory factor (USF) and USF-binding
PT compounds such as amyloid precursor-like proteins APLP1 and APLP2
XX
PS Example 1; Page 78; 130pp; English.
XX
CC 3 Regions (AAR98922-24) of the mouse amyloid precursor-like
CC protein APLP1 (see also AAR98903) show a high degree of homology to
CC the human amyloid precursor protein (APP). The 3 regions
CC respectively correspond to amino acids 21-210, 316-470 and 609-654
CC in extracellular domain I, extracellular domain II and the
CC cytoplasmic domain of APLP1. APLP1 is an upstream stimulatory
CC factor binding protein capable of down-regulating expression
CC from the APP gene promoter. It is a new member of the APP-like
CC family.
XX
SQ Sequence 190 AA;
Query Match 82.5%; Score 33; DB 17; Length 190;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
Db 2 LLPLSLLL 10

RESULT 39
AAB42852
ID AAB42852 standard; Protein: 201 AA.
XX
AC AAB42852;
XX
DT 08-FEB-2001 (first entry)
XX

```

```

Human ORFX ORF2616 polypeptide sequence SEQ ID NO:5232.
DE
XX Human: Open reading frame; ORFX: detection; cytostatic; hepatotropic;
KW vulnery; antipsoriatic; antiparkinsonian; neotropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antifungal; antithyroid;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disease; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN W0200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US08621.
XX
PR 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX
PA (CURA-) CURAGEN CORP.
XX
PI Shinkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
XX
DR N-PSDB; AAC77061.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX
PS Claim 11; Page 4413; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
CC antipsoriatic; antiparkinsonian; neotropic; neuroprotective;
CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
CC antidiabetic; hypotensive; dermatological; antithyroid; antirheumatic;
CC antiinflammatory; antibacterial; antiviral; antifungal; antithyroid;
CC antithyroid; and antianaemic. The sequences can be used for determining
CC the presence of or predisposition to, or preventing or treating
CC pathological conditions associated with an ORFX-associated disorder. The
CC nucleic acids can be used to express ORFX proteins in gene therapy
CC vectors. The proteins and nucleic acids may be used to treat cancers,
CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
CC graft vs host disease, cardiovascular disease, diabetes mellitus,
CC hyperthyroidism, cholesterol ester storage, systemic lupus
CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
CC coagulation; to inhibit thrombosis; and as a contraceptive.
XX
SQ Sequence 201 AA;
Query Match 82.5%; Score 33; DB 21; Length 201;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9
    |||| :|||

```

Db 7 LLPLHLLL 15

RESULT 40
AAB24060

ID AAB24060 standard; Protein; 201 AA.

XX AC AAB24060;

XX DT 29-JAN-2001 (first entry)

XX DE Human PRO535 protein sequence SEQ ID NO:14.

XX KW Human; tumour; diagnosis; neoplastic disease; neoplastic cell growth; proliferation; tumorigenesis; identification; cancer; cytostatic; neurotropic; neuroprotective; antiinflammatory; immunosuppressive; immunostimulant; antiangiogenic; leukaemia; lymphoid malignancy; neuronal disorder; glial disorder; astrocytal disorder; angiogenic; hypothalamic disorder; glandular disorder; macrophagal disorder; epithelial disorder; stromal disorder; blastocoelic disorder; inflammatory disorder; immunologic disorder.

XX KW Homo sapiens.

XX OS WO200053755-A2.

XX PN 14-SEP-2000.

XX PD 06-JAN-2000; 2000WO-US00376.

XX PF 08-MAR-1999; 99WO-US05028.

XX PR 02-JUN-1999; 99WO-US12252.

XX PR 23-JUN-1999; 99US-0141037.

XX PR 07-JUL-1999; 99US-0143048.

XX PR 26-JUL-1999; 99US-0145698.

XX PR 30-NOV-1999; 99WO-US28313.

XX PR 20-DEC-1999; 99WO-US30911.

XX PR 05-JAN-2000; 2000WO-US00219.

XX PA (GETH) GENENTECH INC.

XX PI Ashkenazi AJ, Baker KP, Goddard A, Gurney AL, Hillan KJ, Roy MA; Watanabe CK, Wood WI;

XX PT WPI: 2000-572270/53.

XX DR N-PSDB; AAC58370.

XX PT Thirty PRO polynucleotides encoding PRO polypeptides, useful in the treatment, diagnosis and prevention of cancer -

XX PS Claim 61; Fig 8; 286pp; English.

XX CC The present invention describes an isolated antibody that binds to one of the human PRO proteins designated PRO212, PRO290, PRO341, PRO535, PRO619, PRO717, PRO809, PRO830, PRO848, PRO943, PRO1005, PRO1009, PRO1025, PRO1030, PRO1097, PRO1107, PRO1111, PRO1153, PRO1182, PRO1184, PRO1187, PRO1281, PRO23, PRO39, PRO834, PRO1317, PRO1710, PRO2094, PRO2145 OR PRO2198. PRO antagonists can be used to inhibit tumour cell growth. The PRO polypeptides and nucleotides are useful in the treatment, diagnosis and prevention of cancer. The antibodies and other anti-tumour compounds may be used to treat various conditions, including those characterised by overexpression and/or activation of the amplified PRO genes. Exemplary conditions or disorders to be treated with such antibodies and other compounds include benign or malignant tumours (e.g., renal, liver, kidney, bladder, breast, gastric, ovarian, colorectal, prostate, pancreatic, lung, vulva, thyroid, hepatic carcinomas, sarcomas, glioblastomas, and various head and neck tumours), leukaemias and lymphoid malignancies, other disorders such as neuronal, glial, astrocytal, hypothalamic and other glandular, macrophagal, epithelial, stromal and blastocoelic disorders, and inflammation, angiogenic and immunologic disorders. AAC58242 to AAC58366 represent PCR primers and hybridisation probes used in the isolation of the human PRO sequences. AAC58367 to AAC58396 and AAB24057 to AAB24089 represent human

CC PRO polynucleotide and protein sequences given in the exemplification of the present invention.

XX CC Sequence 201 AA;

XX SQ Query Match 82.5%; Score 33; DB 21; Length 201; Best Local Similarity 77.8%; Pred. No. 1.3e+02; Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LLPLQILL 9

Db 7 LLPLHLLL 15

RESULT 41
AAB23604

ID AAB23604 standard; Protein; 201 AA.

XX AC AAB23604;

XX DT 12-JAN-2001 (first entry)

XX DE Human secreted protein SEQ ID NO: 8.

XX KW Human; secreted protein; cytokine; cell proliferation; nutritional supplement; immune modulation; autoimmune disorder; haematopoiesis regulation; tissue growth; haemostasis; inflammation.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Peptide 7..19

XX FT Protein /label= signal_peptide

XX FT /label= mature_protein

XX PN WO200049134-A1.

XX PD 24-AUG-2000.

XX PF 18-FEB-2000; 2000WO-US04340.

XX PR 19-FEB-1999; 99US-0120680.

XX PR 23-APR-1999; 99US-0298733.

XX PR 17-AUG-1999; 99US-0149639.

XX PR 23-SEP-1999; 99US-0155686.

XX PR 01-OCT-1999; 99US-0157247.

XX PR 29-NOV-1999; 99US-0167822.

XX PR 29-NOV-1999; 99US-0167823.

XX PR 15-FEB-2000; 2000US-0298733.

XX PA (ALPH-) ALPHAGENE INC.

XX PI Valenzuela D, Yuan O, Hoffman H, Hall J, Rapiejko P;

XX DR WPI: 2000-549267/50.

XX DR N-PSDB; AAA93104.

XX PT New secreted proteins and polynucleotides encoding them, which are derived from Homo sapiens, useful for therapy, diagnosis, and research, as well as nutritional sources or supplements -

XX PS Claim 17; Page 243; 309pp; English.

XX CC The present sequence is the sequence of a human secreted protein. Its cDNA was isolated from an adult lung cDNA library. The proteins and coding sequences of the invention can be used in the isolation of similar genes and proteins, in the elucidation of their function in vivo, and to treat a number of conditions. It is possible that they may have uses as nutritional supplements, as cytokine or cell proliferation factors, in immune modulation, where they may be used to treat immune and autoimmune diseases, as haematopoiesis regulators (treating myeloid or lymphoid cell deficiencies), in the promotion of tissue growth, they may

CC have chemokine or chemotactic activity, haemostatic or thrombolytic
CC activity, or anti-inflammatory activity.

Sequence 201 AA;

Query Match 82.5%; Score 33; DB 21; Length 201;

Best Local Similarity 77.8%; Pred. No. 1.3e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LPLQLILL 9

||||:||||

Db 7 LPLHLLLL 15

RESULT 42

AAy66651

ID AAY66651 standard; protein; 201 AA.

AC AAY66651;

XX 05-APR-2000 (first entry)

XX Membrane-bound protein PRO535.

XX Membrane-bound polypeptide; PRO polypeptide; LDL receptor; TIE ligand;
KW pharmaceutical; receptor immunoadhesin; gene mapping.

XX Homo sapiens.

XX WO9963088-A2.

XX 09-DEC-1999.

XX 02-JUN-1999; 99WO-US12252.

XX 02-JUN-1998; 98US-0087607.

PR 02-JUN-1998; 98US-0087609.

PR 02-JUN-1998; 98US-0087759.

PR 03-JUN-1998; 98US-0087827.

PR 04-JUN-1998; 98US-0088021.

PR 04-JUN-1998; 98US-0088025.

PR 04-JUN-1998; 98US-0088028.

PR 04-JUN-1998; 98US-0088029.

PR 04-JUN-1998; 98US-0088030.

PR 04-JUN-1998; 98US-0088033.

PR 04-JUN-1998; 98US-0088326.

PR 05-JUN-1998; 98US-0088157.

PR 05-JUN-1998; 98US-0088202.

PR 05-JUN-1998; 98US-0088212.

PR 05-JUN-1998; 98US-0088217.

PR 09-JUN-1998; 98US-0088655.

PR 10-JUN-1998; 98US-0088722.

PR 10-JUN-1998; 98US-0088730.

PR 10-JUN-1998; 98US-0088734.

PR 10-JUN-1998; 98US-0088738.

PR 10-JUN-1998; 98US-0088740.

PR 10-JUN-1998; 98US-0088741.

PR 10-JUN-1998; 98US-0088742.

PR 10-JUN-1998; 98US-0088810.

PR 10-JUN-1998; 98US-0088811.

PR 10-JUN-1998; 98US-0088824.

PR 10-JUN-1998; 98US-0088825.

PR 10-JUN-1998; 98US-0088826.

PR 11-JUN-1998; 98US-0088858.

PR 11-JUN-1998; 98US-0088861.

PR 11-JUN-1998; 98US-0088863.

PR 11-JUN-1998; 98US-0088876.

PR 12-JUN-1998; 98US-0089090.

PR 12-JUN-1998; 98US-0089105.

PR 16-JUN-1998; 98US-0089440.

PR 16-JUN-1998; 98US-0089512.

PR 16-JUN-1998; 98US-0089514.

PR 17-JUN-1998; 98US-0089532.

PR 17-JUN-1998; 98US-0089538.
PR 17-JUN-1998; 98US-0089598.
PR 17-JUN-1998; 98US-0089599.
PR 17-JUN-1998; 98US-0089600.
PR 17-JUN-1998; 98US-0089653.
PR 18-JUN-1998; 98US-0089801.
PR 18-JUN-1998; 98US-0089907.
PR 18-JUN-1998; 98US-0089908.
PR 19-JUN-1998; 98US-0089947.
PR 19-JUN-1998; 98US-0089948.
PR 19-JUN-1998; 98US-0089952.
PR 22-JUN-1998; 98US-0090246.
PR 22-JUN-1998; 98US-0090252.
PR 22-JUN-1998; 98US-0090254.
PR 23-JUN-1998; 98US-0090349.
PR 23-JUN-1998; 98US-0090355.
PR 24-JUN-1998; 98US-0090429.
PR 24-JUN-1998; 98US-0090431.
PR 24-JUN-1998; 98US-0090435.
PR 24-JUN-1998; 98US-0090444.
PR 24-JUN-1998; 98US-0090445.
PR 24-JUN-1998; 98US-0090461.
PR 24-JUN-1998; 98US-0090472.
PR 24-JUN-1998; 98US-0090535.
PR 24-JUN-1998; 98US-0090538.
PR 24-JUN-1998; 98US-0090540.
PR 24-JUN-1998; 98US-0090557.
PR 25-JUN-1998; 98US-0090676.
PR 25-JUN-1998; 98US-0090678.
PR 25-JUN-1998; 98US-0090688.
PR 25-JUN-1998; 98US-0090690.
PR 25-JUN-1998; 98US-0090691.
PR 25-JUN-1998; 98US-0090694.
PR 25-JUN-1998; 98US-0090695.
PR 25-JUN-1998; 98US-0090696.
PR 26-JUN-1998; 98US-0090862.
PR 26-JUN-1998; 98US-0090863.
PR 01-JUL-1998; 98US-0091358.
PR 01-JUL-1998; 98US-0091360.
PR 01-JUL-1998; 98US-0091544.
PR 02-JUL-1998; 98US-0091478.
PR 02-JUL-1998; 98US-0091486.
PR 02-JUL-1998; 98US-0091519.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091626.
PR 02-JUL-1998; 98US-0091628.
PR 02-JUL-1998; 98US-0091633.
PR 02-JUL-1998; 98US-0091646.
PR 02-JUL-1998; 98US-0091673.
PR 02-JUL-1998; 98US-0091978.
PR 07-JUL-1998; 98US-0091982.
PR 07-JUL-1998; 98US-0092182.
PR 09-JUL-1998; 98US-0092472.
PR 10-JUL-1998; 98US-0093339.
PR 30-JUL-1998; 98US-0094651.
PR 04-AUG-1998; 98US-0095282.
PR 04-AUG-1998; 98US-0095285.
PR 04-AUG-1998; 98US-0095301.
PR 04-AUG-1998; 98US-0095302.
PR 04-AUG-1998; 98US-0095318.
PR 04-AUG-1998; 98US-0095321.
PR 04-AUG-1998; 98US-0095325.
PR 10-AUG-1998; 98US-0095916.
PR 10-AUG-1998; 98US-0095929.
PR 10-AUG-1998; 98US-0096012.
PR 11-AUG-1998; 98US-0096143.
PR 11-AUG-1998; 98US-0096146.
PR 12-AUG-1998; 98US-0096329.
PR 17-AUG-1998; 98US-0096757.
PR 17-AUG-1998; 98US-0096766.
PR 17-AUG-1998; 98US-0096768.
PR 17-AUG-1998; 98US-0096773.
PR 17-AUG-1998; 98US-0096791.
PR 17-AUG-1998; 98US-0096867.


```

PR 17-AUG-1998; 98US-0096891.
PR 17-AUG-1998; 98US-0096894.
PR 17-AUG-1998; 98US-0096895.
PR 17-AUG-1998; 98US-0096897.
PR 18-AUG-1998; 98US-0096949.
PR 18-AUG-1998; 98US-0096950.
PR 18-AUG-1998; 98US-0096959.
PR 18-AUG-1998; 98US-0096960.
PR 18-AUG-1998; 98US-0097022.
PR 19-AUG-1998; 98US-0097141.
PR 20-AUG-1998; 98US-0097218.
PR 24-AUG-1998; 98US-0097661.
PR 26-AUG-1998; 98US-0097951.
PR 26-AUG-1998; 98US-0097952.
PR 26-AUG-1998; 98US-0097954.
PR 26-AUG-1998; 98US-0097955.
PR 26-AUG-1998; 98US-0097971.
PR 26-AUG-1998; 98US-0097974.
PR 26-AUG-1998; 98US-0097978.
PR 26-AUG-1998; 98US-0097979.
PR 26-AUG-1998; 98US-0097986.
PR 26-AUG-1998; 98US-0098014.
PR 31-AUG-1998; 98US-0098525.
PR 16-SEP-1998; 98US-0100634.
PR 12-JAN-1999; 99US-0115565.
XX
PA (GETH ) GENENTECH INC.
XX
XX Baker K, Chen J, Goddard A, Gurney AL, Smith V, Watanabe CK;
PI Wood WI, Yuan J;
XX
XX WPI: 2000-072883/06.
DR N-PSDB; AAF64972.
XX
XX Membrane-bound proteins and related nucleotide sequences -
PT
XX
XX claim 12; Fig 53; 822pp; English.
XX
XX The invention provides membrane-bound PRO polypeptides and
CC polynucleotides encoding them. The PRO sequences of the invention were
CC identified based on extracellular domain homology screening. The PRO
CC sequences have homology with proteins including LDL receptors, TIE
CC ligands and various enzymes. The membrane-bound proteins and receptor
CC molecules are useful as pharmaceutical and diagnostic agents. Receptor
CC immunoadhesins, for instance, can be used as therapeutic agents to block
CC receptor-ligand interactions. The membrane-bound proteins can also be
CC employed for screening of potential peptide or small molecule inhibitors
CC of the relevant receptor/ligand interaction. The PRO encoding sequences
CC are useful as hybridization probes, in chromosome and gene mapping and in
CC the generation of antisense RNA and DNA. PRO nucleic acid sequences
CC will also be useful for the preparation of PRO polypeptides, especially
CC by recombinant techniques.
XX
XX Sequence 201 AA;
XX
XX Query Match 82.5%; Score 33; DB 21; Length 201;
XX Best Local Similarity 77.8%; Pred. No. 1.3e+02;
XX Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 LLLPLQLLL 9
Db 7 LPLHLLLL 15
III :III
LPLHLLLL 15
RESULT 43
AAB65174
ID AAB65174 standard; Protein; 201 AA.
XX
XX AAB65174;
AC
XX
XX 02-APR-2001 (first entry)
DT
XX
XX Human PRO535 (UNQ336) protein sequence SEQ ID NO:99.
DE

```

```

XX Human; secreted and transmembrane protein; PRO; cytostatic;
KW cell death; cancer; chromosomal mapping; gene mapping; tissue typing;
KW diagnostic assay.
XX
XX Homo sapiens.
XX
XX WO200073454-A1.
XX
XX 07-DEC-2000.
XX
XX 30-MAR-2000; 2000WO-US08439.
XX
XX 02-JUN-1999; 99WO-US12252.
PR 23-JUN-1999; 99US-0141037.
PR 07-JUL-1999; 99US-0143048.
PR 20-JUL-1999; 99US-0144758.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 17-AUG-1999; 99US-0149396.
PR 15-SEP-1999; 99WO-US21090.
PR 15-SEP-1999; 99WO-US21547.
PR 08-OCT-1999; 99US-0158663.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28301.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 05-JAN-2000; 2000WO-US00219.
PR 06-JAN-2000; 2000WO-US00376.
PR 11-FEB-2000; 2000WO-US03565.
PR 18-FEB-2000; 2000WO-US04341.
PR 22-FEB-2000; 2000WO-US04414.
PR 24-FEB-2000; 2000WO-US04914.
PR 24-FEB-2000; 2000WO-US05004.
PR 02-MAR-2000; 2000WO-US05841.
PR 15-MAR-2000; 2000WO-US06884.
PR 20-MAR-2000; 2000WO-US07377.
XX
XX (GETH ) GENENTECH INC.
XX
XX Ashkenazi AJ, Baker KP, Botstein D, Desnoyers L, Eaton DL;
PI Ferrara N, Fong S, Gerber H, Gerritsen ME, Goddard A, Godowski PJ;
PI Grimaldi CJ, Gurney AL, Kljavin IJ, Napier MA, Pan J, Paoni NF;
PI Roy MA, Stewart TA, Tumas D, Watanabe CK, Williams PM, Wood WI;
PI Zhang Z;
XX
XX WPI: 2001-032160/04.
DR N-PSDB; AAF44118.
XX
XX PRO polynucleotides used to produce polypeptides used to target
PT bioactive molecules such as toxins, radiolabels or antibodies, to
PT specific cells, to cause targeted cell death -
XX
XX Claim 12; Fig 53; 935pp; English.
XX
XX The present invention describes human secreted and transmembrane PRO
CC proteins. The PRO proteins have cytostatic activity. The PRO proteins
CC can be used for targeted delivery of bioactive molecules, such as
CC toxins, radiolabels or antibodies, that cause cell death. PRO nucleotide
CC sequences, and their fragments, can be used as hybridisation probes, in
CC chromosomal and gene mapping, and in the generation of anti-sense RNA
CC and DNA. They may also be used to produce transgenic animals which are
CC used to develop and screen therapeutically useful reagents. The PRO
CC nucleotide and protein sequence can be used for tissue typing and in
CC treating cancer. Anti-PRO antibodies can be used in diagnostic assays.
CC AAF44270 to AAF44470 represent PCR primers and hybridisation probes used
CC in the isolation of human PRO sequences. AAF44087 to AAF44269 and
CC AAB65154 to AAB65300 represent human PRO polynucleotide and protein
CC sequences given in the exemplification of the present invention.
XX
XX Sequence 201 AA;
XX
XX Query Match 82.5%; Score 33; DB 22; Length 201;

```

Best Local Similarity 77.8%; Pred. No. 1.3e+02; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9
| | | | : | | |
Db 7 LLPLHLLL 15

RESULT 44

AAB50984
ID AAB50984 standard; Protein; 201 AA.

XX AAB50984;

XX 21-MAR-2001 (first entry)

DE Human PRO535 protein.

XX Human; PRO; cardiant; antiangiogenic; antiarteriosclerotic; hypotensive;
KW vasotropic; antirheumatic; antiarthritic; antiinflammatory; cytostatic;
KW vulnary; antiangular; gene therapy; cardiovascular disease;
KW endothelial disorder; angiogenic disorder; cancer; periodontal disease;
KW wound healing.

XX Homo sapiens.

XX WO200073445-A2.

XX 07-DEC-2000.

XX 17-MAY-2000; 2000WO-US13705.

XX 02-JUN-1999; 99WO-US12252.

XX 23-JUN-1999; 99US-0141037.

XX 26-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 28-JUL-1999; 99US-0146222.

XX 01-SEP-1999; 99WO-US20111.

XX 30-NOV-1999; 99WO-US28313.

XX 30-NOV-1999; 99WO-US28409.

XX 02-DEC-1999; 99WO-US28565.

XX 16-DEC-1999; 99WO-US30095.

XX 11-FEB-2000; 2000WO-US03565.

XX 18-FEB-2000; 2000WO-US04341.

XX 18-FEB-2000; 2000WO-US04342.

XX 24-FEB-2000; 2000WO-US05004.

XX 02-MAR-2000; 2000WO-US05841.

XX 10-MAR-2000; 2000WO-US06319.

XX 15-MAR-2000; 2000WO-US06884.

XX 21-MAR-2000; 2000WO-US07532.

XX 30-MAR-2000; 2000WO-US08439.

XX (GETH) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Gerritsen ME;

PI Goddard A, Godowski PJ, Gurney AL, Kuo SS, Mark MR, Marsters SA;

PI Paoni NF, Pitti RM, Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-025251/03.

XX N-PSDB; AAC90568.

XX Seventeen nucleic acids encoding PRO polypeptides which are useful in
PT diagnosis and treatment of cardiovascular, endothelial or angiogenic
PT disorders in a mammal -

XX Claim 71; Fig 12; 182pp; English.

XX The present sequence is one of seventeen novel PRO polypeptides. The PRO
CC nucleic acids, polypeptides, agonists and antagonists are useful for
CC treating cardiovascular, endothelial or angiogenic disorders in a mammal.
CC Examples of these disorders include cardiac hypertrophy, trauma, cancer,
CC age-related macular degeneration, atherosclerosis, hypertension, arterial
CC restenosis, Reynaud's disease, rheumatoid arthritis, angina, myocardial

CC infarctions, thrombophlebitis and lymphangitis. The PRO polypeptides and
CC antagonists are also used to prevent tumour angiogenesis and for treating
CC periodontal diseases. They are also used to stimulate wound healing and
CC tissue regeneration. The PRO nucleic acids, polypeptides and anti-PRO
CC antibodies are useful for diagnosing a cardiovascular, endothelial or
CC angiogenic disorder.

XX Sequence 201 AA;

Query Match 82.5%; Score 33; DB 22; Length 201;

Best Local Similarity 77.8%; Pred. No. 1.3e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLPLQILL 9

Db 7 LLPLHLLL 15

RESULT 45

AAB53092

ID AAB53092 standard; Protein; 201 AA.

XX AAB53092;

XX 28-FEB-2001 (first entry)

XX Human angiogenesis-associated protein PRO535, SEQ ID NO:154.

XX Human; angiogenesis-associated protein; PRO; endothelial cell growth;
KW cardiac hypertrophy; cardiovascular disorder; endothelial disorder;
KW angiogenic disorder; atherosclerosis; osteoporosis; hypertension;
KW myocardial infarction; diabetic retinopathy; rheumatoid arthritis;
KW Crohn's disease; psoriasis; endometriosis; ulcer; wound healing; cancer;
KW Alzheimer's disease; Huntington's disease; stroke; drug screening;
KW gene therapy; transgenic animal.

XX Homo sapiens.

XX WO2000053753-A2.

XX 14-SEP-2000.

XX 05-JAN-2000; 2000WO-US00219.

XX 08-MAR-1999; 99WO-US05028.

XX 12-MAR-1999; 99US-0123957.

XX 14-MAY-1999; 99US-0134287.

XX 02-JUN-1999; 99WO-US12252.

XX 23-JUN-1999; 99US-0141037.

XX 20-JUL-1999; 99US-0144758.

XX 26-JUL-1999; 99US-0145698.

XX 01-SEP-1999; 99WO-US20111.

XX 08-SEP-1999; 99WO-US20594.

XX 15-SEP-1999; 99WO-US21090.

XX 15-SEP-1999; 99WO-US21547.

XX 05-OCT-1999; 99WO-US23089.

XX 30-NOV-1999; 99WO-US28313.

XX 30-NOV-1999; 99WO-US28409.

XX 02-DEC-1999; 99WO-US28564.

XX 02-DEC-1999; 99WO-US28565.

XX (GETH) GENENTECH INC.

XX Ashkenazi AJ, Baker KP, Ferrara N, Gerber H, Goddard A;

PI Godowski PJ, Gurney AL, Hillan KJ, Kuo SS, Mark MR, Marsters SA;

PI Paoni NF, Pitti RM, Watanabe CK, Williams PM, Wood WI;

XX WPI; 2001-090793/10.

XX N-PSDB; AAC97489.

XX New isolated nucleic acid for producing a PRO polypeptide, analyzing
PT genetic disorders and treating cardiovascular, endothelial or
PT angiogenic disorders, such as atherosclerosis, wounds or cancer -

XX

PS Claim 69; Fig 58; 293pp; English.

XX

The invention relates to novel human angiogenesis-associated proteins designated PRO proteins (AAB53064-B53097), and to nucleic acids encoding PRO proteins. The invention also relates to vectors and host cells comprising a PRO nucleic acid, the recombinant production of a PRO protein, PRO antibodies specific for a PRO protein, fusion proteins comprising a PRO protein, agonists or antagonists of a PRO protein, and compounds which inhibit the expression of a PRO gene. The invention additionally encompasses methods of identifying modulators of PRO expression or activity; diagnosing a cardiovascular, endothelial or angiogenic disorder, or a susceptibility to such a disorder by detecting mutations in a PRO gene, or the expression level of a PRO gene within a particular tissue; treating a cardiovascular, endothelial or angiogenic disorder via the administration of a PRO protein, PRO nucleic acid, or PRO agonist or antagonist; a retroviral gene therapy vector comprising a PRO nucleic acid; and methods of inhibiting or stimulating endothelial cell growth, cardiac hypertrophy or PRO-induced angiogenesis via the administration of a PRO protein, or an agonist or antagonist thereof. PRO nucleic acids, PRO proteins, antibodies against PRO proteins, PRO agonists and PRO antagonists may be used as therapeutic agents to treat cardiovascular, endothelial or angiogenic disorders, such as atherosclerosis, osteoporosis, myocardial infarction, hypertension, diabetic retinopathy, rheumatoid arthritis, Crohn's disease, psoriasis, endometriosis, ulcers, wounds, cancer, Alzheimer's disease, Huntington's disease, or stroke. PRO nucleic acids are additionally useful in the recombinant production of PRO proteins, as hybridisation probes to screen libraries to isolate cDNAs with sequence identity to PRO proteins, to map genes encoding PRO proteins, to analyse genetic disorders, and in gene therapy. PRO nucleic acids can also be used to produce transgenic animals useful for the development and screening of potential therapeutic agents. The present sequence represents a PRO protein of the invention.

Sequence 201 AA:

Query Match 82.5%; Score 33; DB 22; Length 201;

Best Local Similarity 77.8%; Pred. No. 1.3e+02;

Best Local Matches	Similarity	Prod. No.	Indels	Gaps
7;	77.08;	1; 3502;	0;	0;
Conservative	1;	Mismatches	1;	Indels
7;	Conservative	1;	Mismatches	1;
Indels	0;	Gaps	0;	Gaps

Qy 1 LLPLQILL 9

		:	
		:	
		:	
		:	

Search completed: November 6, 2002, 12:05:10

Job time : 25.7778 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 : Search time 18.4444 seconds
(without alignments)
84.413 Million cell updates/sec

Title: US-09-905-083-36
Perfect score: 40
Sequence: 1 SLLPLQLI 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL19:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertebrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriaph:*
17: sp.archaea:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	95.0	146	17 Q9V2D5	Q9V2d5 pyrococcus
2	37	92.5	208	8 O21527	O21527 clethrionom
3	34	85.0	189	11 Q9D8P8	Q9D8p8 mus musculus
4	34	85.0	198	11 Q9CRE4	Q9cre4 mus musculus
5	34	85.0	201	11 Q9D1M7	Q9dlm7 mus musculus
6	33	82.5	1017	11 Q99K41	Q99k41 mus musculus
7	32	80.0	196	11 Q9D226	Q9d226 mus musculus
8	32	80.0	218	10 Q9ZTE3	Q9zte3 arabidopsis
9	32	80.0	256	11 Q9Z2G5	Q9z2g5 mus musculus
10	32	80.0	266	11 Q9ESS3	Q9ess3 mus musculus
11	32	80.0	267	11 Q9RLS4	Q9rls4 rattus norv
12	32	80.0	267	11 O35426	O35426 mus musculus
13	32	80.0	282	10 Q9M0J5	Q9m0j5 arabidopsis
14	32	80.0	404	2 O86166	O86166 porphyromon
15	32	80.0	622	3 Q12212	Q12212 saccharomyc
16	32	80.0	662	5 Q969D0	Q969d0 giardia lam

17	32	80.0	1105	10 Q9FL49	Q9fl49 arabidopsis
18	31	77.5	117	4 Q9PLI9	Q9pli9 homo sapien
19	31	77.5	189	5 Q9VYN0	Q9vyn0 drosophila
20	31	77.5	196	2 Q9X620	Q9x620 salmonella
21	31	77.5	245	11 Q9JMG3	Q9jmg3 mus musculus
22	31	77.5	246	4 Q9BVT8	Q9bvt8 homo sapien
23	31	77.5	247	5 O76155	O76155 periplaneta
24	31	77.5	303	10 O22910	O22910 arabidopsis
25	31	77.5	316	5 Q9V3Y4	Q9v3y4 drosophila
26	31	77.5	347	8 Q9MMS8	Q9mms8 syrigma sib
27	31	77.5	392	16 Q9KAL1	Q9kall bacillus ha
28	31	77.5	395	16 P96675	P96675 bacillus su
29	31	77.5	401	8 Q98SC8	Q98sc8 guillardia
30	31	77.5	438	5 O44312	O44312 drosophila
31	31	77.5	469	2 Q93I52	Q93i52 sphaerofilu
32	31	77.5	470	8 Q9SEA5	Q9seas guillardia
33	31	77.5	475	17 O26479	O26479 methanother
34	31	77.5	519	2 Q9S335	Q9s335 prochloroco
35	31	77.5	630	5 Q95Y88	Q95y88 caenorhabdi
36	31	77.5	934	13 Q9DER4	Q9der4 gallus gall
37	31	77.5	934	13 Q910B7	Q910b7 coturnix co
38	30	75.0	69	6 Q9GJR4	Q9gjr4 ovis aries
39	30	75.0	82	15 Q9WI77	Q9wi77 human immun
40	30	75.0	121	11 Q9DOC8	Q9doc8 mus musculus
41	30	75.0	150	16 Q9CKE2	Q9cke2 pasteurella
42	30	75.0	155	6 Q9GL83	Q9gl83 capra hircu
43	30	75.0	155	6 Q95MP4	Q95mp4 ovis aries
44	30	75.0	155	6 Q95RP3	Q95rp3 bubalus bub
45	30	75.0	162	12 Q91T40	Q91t40 lumpy skin

ALIGNMENTS

RESULT 1

Q9V2D5 PRELIMINARY; PRT; 146 AA.
AC Q9V2D5;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE HYPOTHETICAL 16.1 KDA PROTEIN.
GN PAB0088.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=29292;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ORSAY;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ248283; CAB49063.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 146 AA; 16092 MW; 7182941371258C1F CRC64;

Query Match 95.0%; Score 38; DB 17; Length 146;
Best Local Similarity 88.9%; Pred. No. 2.4;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQLI 9

Db 66 SLLPLQLII 74

RESULT 2

O21527 PRELIMINARY; PRT; 208 AA.
ID O21527
AC O21527;
DT 01-JAN-1998 (Tremblrel. 05, Created)
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)

```

DE - ND4.
GN NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3) (FRAGMENT).
OS Clethrionomys gapperi (Southern red-backed vole).
OC Mitochondrion
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Arvicolinae;
OC Clethrionomys.
OX NCBI_TaxID=56223;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98152303; PubMed=9491603;
RA Engel S.R., Hogan K.M., Taylor J.F., Davis S.K.;
RT "Molecular systematics and paleobiogeography of the South American
RL sigmodontine rodents.";
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL: U83808; AAB87168.1; -.
DR InterPro: IPR001750; OXidored_q1.
DR InterPro: IPR000260; OXidored_q5_N.
DR Pfam: PF00361; OXidored_q1; 1.
DR Pfam: PF01059; OXidored_q5_N; 1.
KW Mitochondrion; NAD; OXidoreductase; Ubiquinone.
FT NON_TER 208
SQ SEQUENCE 208 AA; 23967 MW; 8AF1788697AED6A2 CRC64;

Query Match 92.5%; Score 37; DB 8; Length 208;
Best Local Similarity 88.9%; Pred. No. 5.5;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 97 SLLPLQL 105

RESULT 3
Q9D8P8 PRELIMINARY; PRT; 189 AA.
AC Q9D8P8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE I110002023RIK PROTEIN.
GN I110002023RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=PANCREAS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
EMBL: AK007813; BAB25278.1; -.
HSSP: P20071; 1FKJ.
DR MGD; MGI:1913370; I110002023RIK.
DR InterPro: IPR001179; FKBP_PPIase.
DR PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
DR PROSITE; PS00509; FKBP_PPIASE_3; 1.
FT NON_TER 198
SQ SEQUENCE 198 AA; 21882 MW; A588345383032972 CRC64;

Query Match 85.0%; Score 34; DB 11; Length 198;
Best Local Similarity 87.5%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
Db 6 LLLPLQL 13

RESULT 5
Q9D1M7 PRELIMINARY; PRT; 198 AA.
AC Q9D1M7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE I110002023RIK PROTEIN (FRAGMENT).
GN I110002023RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
EMBL: AK007813; BAB25278.1; -.
HSSP: P20071; 1FKJ.

```

```

DR MGD; MGI:1913370; I110002023RIK.
DR InterPro: IPR001179; FKBP_PPIase.
DR Pfam: PF00254; FKBP; 1.
DR PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
DR PROSITE; PS00509; FKBP_PPIASE_3; 1.
SQ SEQUENCE 189 AA; 20626 MW; AD9795B7F1E0582B CRC64;

Query Match 85.0%; Score 34; DB 11; Length 189;
Best Local Similarity 87.5%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
Db 6 LLLPLQL 13

RESULT 4
Q9CRE4 PRELIMINARY; PRT; 198 AA.
AC Q9CRE4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE I110002023RIK PROTEIN (FRAGMENT).
GN I110002023RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
EMBL: AK019132; BAB31559.1; -.
HSSP: P20071; 1FKJ.
DR MGD; MGI:1913370; I110002023RIK.
DR InterPro: IPR001179; FKBP_PPIase.
DR PROSITE; PS00453; FKBP_PPIASE_1; UNKNOWN_1.
DR PROSITE; PS00509; FKBP_PPIASE_3; 1.
FT NON_TER 198
SQ SEQUENCE 198 AA; 21882 MW; A588345383032972 CRC64;

Query Match 85.0%; Score 34; DB 11; Length 198;
Best Local Similarity 87.5%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
Db 6 LLLPLQL 13

RESULT 5
Q9D1M7 PRELIMINARY; PRT; 198 AA.
AC Q9D1M7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE I110002023RIK PROTEIN (FRAGMENT).
GN I110002023RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=EMBRYONIC STEM CELLS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
EMBL: AK019132; BAB31559.1; -.
HSSP: P20071; 1FKJ.

```

```
ID Q9D1M7 PRELIMINARY; PRT; 201 AA.
AC Q9D1M7;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 1110002023RIK PROTEIN.
GN 1110002023RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaio I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK003331; BAB22719.1; -.
DR HSSP; P20071; 1FKJ.
DR MGD; MGI:1913370; 1110002023RIK.
DR InterPro; IPR001179; FKBP_PP1ase.
DR PROSITE; PS00453; FKBP_PP1ase_1; UNKNOWN_1.
DR PROSITE; PS00599; FKBP_PP1ase_3; 1.
SQ SEQUENCE 201 AA; 22137 MW; 94D955C57264BD82 CRC64;

Query Match 85.0%; Score 34; DB 11; Length 201;
Best Local Similarity 87.5%; Pred. No. 23;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
   |||||:|
DB 6 LLLPLQLL 13

RESULT 6
Q99K41 PRELIMINARY; PRT; 1017 AA.
AC Q99K41;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SIMILAR TO ELASTIN MICROFIBRIL INTERFACE LOCATED PROTEIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=MAMMARY TUMOR. WAP-TGF ALPHA MODEL. 7 MONTHS OLD, GROSS
RC TISSUE.;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC005481; AAH05481.1; -.
DR InterPro; IPR001073; Clq.
DR InterPro; IPR000087; Collagen.
```

```
DR Pfam; PF00386; Clq; 1.
DR Pfam; PF01391; Collagen; 1.
DR SMART; SM00110; Clq; 1.
SQ SEQUENCE 1017 AA; 107584 MW; 2EF903204DB9C88F CRC64;

Query Match 82.5%; Score 33; DB 11; Length 1017;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLLPLOI 8
   |||||:|
DB 973 SLLPLQV 980

RESULT 7
Q9D226 PRELIMINARY; PRT; 196 AA.
ID Q9D226;
AC Q9D226;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE A030007E19RIK PROTEIN.
GN A030007E19RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SKIN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaio I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK020699; BAB32180.1; -.
DR HSSP; P10969; 1WGT.
DR MGD; MGI:1924465; A030007E19RIK.
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; UNKNOWN_1.
SQ SEQUENCE 196 AA; 17777 MW; 2136738FD496D815 CRC64;

Query Match 80.0%; Score 32; DB 11; Length 196;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLLPLOIL 9
   |||||:|
DB 145 ALLLPVQLL 153

RESULT 8
Q9ZTE3 PRELIMINARY; PRT; 218 AA.
ID Q9ZTE3;
AC Q9ZTE3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
```

01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE PUTATIVE TRANSCRIPTION FACTOR (FRAGMENT).
GN MYB41 OR T13J8.220.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=99056848; PubMed=9839469;
RA Kranz H.D., Denekamp M., Greco R., Jin H.-L., Leyva A., Weissner R.,
RA Petroni K., Urzainqui A., Bevan M., Martin C., Smeekens S.,
RA Tonelli C., Paz-Ares J., Weisshaar B.,
RT "Towards functional characterisation of the members of the R2R3-MYB
gene family from Arabidopsis thaliana.";
RL Plant J. 16:263-276(1998).
RN [2]
RP SEQUENCE FROM N.A.
RA Bevan M., Pohl T., Weizenegger T., Hohelsel J., Mewes H.W.,
RA Mayer K.F.X., Schueller C.,
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF062882; RAC83604.1; -;
DR EMBL: AL035524; CAB36780.1; -;
DR HSSP: P01103; IPOM.
DR InterPro: IPR001005; Myb_DNA_bind.
DR Pfam: PF00249; myb_DNA-binding: 1.
DR SMART: SM00395; SANT: 1.
DR PROSITE: PS00334; MYB_2: 1.
DR PROSITE: PS50090; MYB_3: 1.
FT NON_TER 1
SQ SEQUENCE 218 AA; 24240 MW; D783569F1C8F215E CRC64;

Query Match 80.0%; Score 32; DB 10; Length 218;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQ 7
DB 104 SLLPLQ 110
|||||

RESULT 9
Q922G5 ID Q922G5 PRELIMINARY; PRT; 256 AA.
AC Q922G5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE SIMILAR TO X-BOX BINDING PROTEIN 1 (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC008153; AAH08153.1; -;
FT NON_TER 1
SQ SEQUENCE 256 AA; 28651 MW; 5544DB566D782620 CRC64;

Query Match 80.0%; Score 32; DB 11; Length 256;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQL 9
DB 187 LLPLQL 193
|||||

RESULT 11
Q9R1S4 ID Q9R1S4 PRELIMINARY; PRT; 267 AA.
AC Q9R1S4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HEPATOCARCINOGENESIS-RELATED TRANSCRIPTION FACTOR (HTF).
GN HTF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=20137507; PubMed=10675042;
RA Kokura K., Kishimoto T., Tamura T.;
RT "Identity between rat htf and human xbp-1 genes: determination of gene
structure, target sequence, and transcription promotion function for
HTF.";
RL Gene 241:297-307(2000).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL: AB030238; BAA82600.1; -;
DR InterPro: IPR001871; bZIP.
DR Pfam: PF00170; bZIP: 1.
DR SMART: SM00338; BRLZ: 1.
DR PROSITE: PS00036; BZIP_BASIC: 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 267 AA; 29665 MW; B5A58F1D3FAA10B4 CRC64;

Db 177 LLPLQL 183
RESULT 10
Q9ESS3 ID Q9ESS3 PRELIMINARY; PRT; 266 AA.
AC Q9ESS3;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE TAX-RESPONSIVE ELEMENT-BINDING PROTEIN 5.
GN TREB5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20363095; PubMed=10907849;
RA Masaki T., Noguchi H., Kobayashi M., Yoshida M., Takamatsu K.;
RT "Isolation and characterization of the gene encoding mouse tax-
responsive element-binding protein (TREB)5.";
RL DNA Res. 7:187-193(2000).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL: AB036745; BAB13793.1; -;
DR InterPro: IPR001871; bZIP.
DR SMART: SM00338; BRLZ: 1.
DR PROSITE: PS00036; BZIP_BASIC: 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 266 AA; 29562 MW; 4161FFA93B1800A5 CRC64;

Query Match 80.0%; Score 32; DB 11; Length 266;
Best Local Similarity 100.0%; Pred. No. 76;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQL 9
DB 187 LLPLQL 193
|||||

RESULT 11
Q9R1S4 ID Q9R1S4 PRELIMINARY; PRT; 267 AA.
AC Q9R1S4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HEPATOCARCINOGENESIS-RELATED TRANSCRIPTION FACTOR (HTF).
GN HTF.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR;
RX MEDLINE=20137507; PubMed=10675042;
RA Kokura K., Kishimoto T., Tamura T.;
RT "Identity between rat htf and human xbp-1 genes: determination of gene
structure, target sequence, and transcription promotion function for
HTF.";
RL Gene 241:297-307(2000).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.
DR EMBL: AB030238; BAA82600.1; -;
DR InterPro: IPR001871; bZIP.
DR Pfam: PF00170; bZIP: 1.
DR SMART: SM00338; BRLZ: 1.
DR PROSITE: PS00036; BZIP_BASIC: 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 267 AA; 29665 MW; B5A58F1D3FAA10B4 CRC64;

Query Match 80.0%; Score 32; DB 11; Length 267;
Best Local Similarity 100.0%; Pred. No. 79;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQIL 9

|||||||

Db 188 LLPLQIL 194

RESULT 12

O35426 PRELIMINARY; PRT; 267 AA.

AC O35426;

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE X BOX BINDING PROTEIN-1.

GN XBP1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RA Lee C.M., Reddy E.P.;

RT "Sequence Analysis of Murine XBP-1.";

RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE BZIP FAMILY.

DR EMBL; AF027963; AAB81862.2; -

DR MGD; MGI:98970; XBP1.

DR InterPro: IPR001871; bZIP.

DR SMART; SM00338; BRLZ; 1.

DR PROSITE; PS00036; BZIP_BASIC; 1.

DR DNA-binding; Nuclear protein.

QY SEQUENCE 267 AA; 29619 MW; 15009E684F6D426F CRC64;

Query Match

Best Local Similarity 80.0%; Score 32; DB 11; Length 267;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQIL 9

|||||||

Db 188 LLPLQIL 194

RESULT 13

O9M0J5 PRELIMINARY; PRT; 282 AA.

AC O9M0J5;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)

DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE PUTATIVE TRANSCRIPTION FACTOR MYB41.

GN AT4G28110.

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Pohl T., Weizenegger T., Meves H.W., Lemcke K., Mayer K.F.X.;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL161572; CAB79613.1; -

DR HSSP; P01103; IPOM.

DR InterPro: IPR001005; Myb_DNA_bind.

DR Pfam; PF00249; myb_DNA-binding; 2.

DR SMART; SM00395; SANT; 2.

DR PROSITE; PS00037; MYB_1; 1.
DR PROSITE; PS00334; MYB_2; 1.
DR PROSITE; PS50090; MYB_3; 2.
SQ SEQUENCE 282 AA; 31651 MW; 085D065C7980F7FE CRC64;

Query Match 80.0%; Score 32; DB 10; Length 282;
Best Local Similarity 100.0%; Pred. No. 83;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQ 7

|||||||

Db 168 SLLPLQ 174

RESULT 14

O86166 PRELIMINARY; PRT; 404 AA.

AC O86166;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE PORS PROTEIN.

GN PORS.

OS Porphyromonas gingivalis (Bacteroides gingivalis).

OC Bacteria; CFB group; Bacteroidetes; Bacteroidales; Porphyromonadaceae;

OC Porphyromonas.

OX NCBI_TaxID=837;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC33277;

RA Nakayama K., Ratnayake D.B.;

RT "Identification, cloning and sequencing of a gene (pOR) responsible

for black pigmentation and extracellular protease production of

Porphyromonas gingivalis.";

RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=ATCC33277;

RA Nakayama K.;

RT "porr and pors genes of Porphyromonas gingivalis.";

RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.

DR EMBL; D64132; BAA31965.1; -

DR InterPro: IPR002088; PPTA.

DR PROSITE; PS00904; PPTA; UNKNOWN_1.

SQ SEQUENCE 404 AA; 45748 MW; 3DA49FE2239EBEEA CRC64;

Query Match 80.0%; Score 32; DB 2; Length 404;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQ 7

|||||||

Db 215 SLLPLQ 221

RESULT 15

Q12212 PRELIMINARY; PRT; 622 AA.

AC Q12212;

DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE YOR3329C FROM CHROMOSOME XV.

GN YOR137C OR YOR3329C.

OS Saccharomyces cerevisiae (Baker's yeast).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.

OX NCBI_TaxID=4932;

RN [1]

RP SEQUENCE FROM N.A.

RA Benes V., Andrade M.A., Rechmann S., Teodoru C., Banreivi A.,

RA Sander C., Valencia A., Ansoorge W., Voss H.;

RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

```

RA  SEQUENCE FROM N.A.
RA  Voss H., Benes V., Rechmann S., Teodoru C., Schwager C., Paces V.,
RA  Ansoerge W.;
RA  Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RA  [3]
RA  SEQUENCE FROM N.A.
RA  MTPS;
RA  Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RA  EMBL: X94335; CAA64055.1; -.
RA  EMBL: 275045; CAA99336.1; -.
RA  SGD: S0005663; YOR137C.
RA  InterPro: IPR000934; Ser_thr_phosphptse.
RA  SEQUENCE 622 AA; 71845 MW; 8DE06835FE88FB14 CRC64;

Query Match      80.0%; Score 32; DB 3; Length 622;
Best Local Similarity 56.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY  1 SLLPLQIL 9
   | :|:|:|
Db  141 SALIPLQVL 149.

RESULT 16
ID  Q969D0      PRELIMINARY;      PRT;      662 AA.
AC  Q969D0;
DT  01-DEC-2001 (TEMBLrel. 19, Created)
DT  01-DEC-2001 (TEMBLrel. 19, Last sequence update)
DT  01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE  KINESIN-LIKE PROTEIN 1 (FRAGMENT).
GN  GLKLP1.
OS  Giardia lamblia (Giardia intestinalis).
OC  Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX  NCBI_TaxID=5741;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=POLISH;
RA  Iwabe N.;
RT  "Giardia lamblia kinesin-like protein.";
RL  Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AB028051; BAB56138.1; -.
FT  NON_TER
   1
SQ  SEQUENCE 662 AA; 72349 MW; 5D215D8F9634213A CRC64;

Query Match      80.0%; Score 32; DB 5; Length 562;
Best Local Similarity 100.0%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  2 LLLPLQI 8
   | :|:|:|
Db  82 LLLPLQI 88

RESULT 17
ID  Q9FL49      PRELIMINARY;      PRT;      1105 AA.
AC  Q9FL49;
DT  01-MAR-2001 (TEMBLrel. 16, Created)
DT  01-MAR-2001 (TEMBLrel. 16, Last sequence update)
DT  01-MAR-2001 (TEMBLrel. 16, Last annotation update)
DE  SIMILARITY TO UNKNOWN PROTEIN.
OS  Arabidopsis thaliana (Mouse-ear cress).
OC  Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC  Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC  eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX  NCBI_TaxID=3702;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=COLUMBIA;
RA  MEDLINE=98344145; PubMed=9679202;
RA  Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
```

```

RA  Tabata S.;
RA  "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
RA  features of the regions of 1,381,565 bp covered by twenty one
RA  physically assigned p1 and TAC clones.";
RA  DNA Res. 5:131-145(1998).
DR  EMBL: AB010697; BAB11155.1; -.
SQ  SEQUENCE 1105 AA; 124725 MW; B6BF4684872FD7D2 CRC64;

Query Match      80.0%; Score 32; DB 10; Length 1105;
Best Local Similarity 66.7%; Pred. No. 3.1e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY  1 SLLPLQIL 9
   | :|:|:|
Db  328 SLLVPLELL 336

RESULT 18
ID  Q9PLI9      PRELIMINARY;      PRT;      117 AA.
AC  Q9PLI9;
DT  01-OCT-2000 (TEMBLrel. 15, Created)
DT  01-OCT-2000 (TEMBLrel. 15, Last sequence update)
DT  01-OCT-2000 (TEMBLrel. 15, Last annotation update)
DE  PRO1618.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=LIVER;
RA  Zhang C., Yu Y., Zhang S., Wei H., Zhou G., Ouyang S., Luo L., Bi J.,
RA  Liu M., He F.;
RT  "Functional prediction of the coding sequences of 121 new genes
RT  deduced by analysis of cDNA clones from human fetal liver.";
RL  Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR  EMBL: AF116642; AAF71065.1; -.
SQ  SEQUENCE 117 AA; 13457 MW; C59D47ADCDC60C23 CRC64;

Query Match      77.5%; Score 31; DB 4; Length 117;
Best Local Similarity 75.0%; Pred. No. 58;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY  1 SLLPLQI 8
   | :|:|:|
Db  79 ALLPLQL 86

RESULT 19
ID  Q9VYN0      PRELIMINARY;      PRT;      189 AA.
AC  Q9VYN0;
DT  01-MAY-2000 (TEMBLrel. 13, Created)
DT  01-MAY-2000 (TEMBLrel. 13, Last sequence update)
DT  01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE  CG15927 PROTEIN.
GN  CG15927.
OS  Drosophila melanogaster (Fruit fly).
OC  Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC  Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC  Ephydroidea; Drosophilidae; Drosophila.
OX  NCBI_TaxID=7227;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=BERKELEY;
RA  Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA  Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA  George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA  Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA  Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA  Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
```

RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Folsler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jallali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RT Science 287:2185-2195(2000).
 DR ENBL: AE003489; AAF48162.1; -;
 DR FlyBase; FBgn0030389; CG15927.
 SQ SEQUENCE 189 AA; 22107 MW; 823E2AF6F3F4BF CRC64;

Query Match 77.5%; Score 31; DB 5; Length 189;
 Best Local Similarity 87.5%; Pred. No. 91;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQIL 9
 |||||
 Db 42 LLLPLNIL 49

RESULT 20
 Q9X620
 ID Q9X620 PRELIMINARY; PRT; 196 AA.
 AC
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CORF.
 GN CORF.
 OS *Salmonella typhimurium*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Salmonella*.
 OC *Salmonella*.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=L72;
 RA Smith R.L., Ahuga D., Thacker L.K., Maguire M.E.;
 RT "Magnesium transport in *Salmonella typhimurium*: Sequence and
 RL characterization of the corB, corC, and corD genes.";
 DR Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
 DR ENBL: AF130857; AAD31438.1; -;
 DR InterPro: IPR002550; DUF21.
 DR Pfam: PF01595; DUF21; 1.
 SQ SEQUENCE 196 AA; 21859 MW; B06659F7E5BAA17A CRC64;

Query Match 77.5%; Score 31; DB 2; Length 196;
 Best Local Similarity 87.5%; Pred. No. 95;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQIL 9
 |||||
 Db 123 LLAPLQIL 130

RESULT 21
 Q9JMG3
 ID Q9JMG3 PRELIMINARY; PRT; 245 AA.
 AC Q9JMG3
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE MRNA, COMPLETE CDS, CLONE:1-2 (2010004020RIK PROTEIN).
 GN AB030183 OR 2010004020RIK.
 OS *Mus musculus* (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Musinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=WHOLE BRAIN;
 RX MEDLINE=20145471; PubMed=10679242;
 RA Inoue S., Sano H., Ohta M.;
 RT "Growth suppression of *Escherichia coli* by induction of expression of
 RL mammalian genes with transmembrane or ATPase domains.";
 RL Biochem. Biophys. Res. Commun. 268:553-561(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=SMALL INTESTINE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Azawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Stauber F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaudo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 DR ENBL: AB030183; BAA92747.1; -;
 DR EMBL: AK008109; BAB25465.1; -;
 DR MGD: MGI:1923764; 2010004020RIK.
 DR InterPro: IPR000626; Ubiquitin.
 DR SMART: SM00213; UBO; 1.
 DR PROSITE: PS00553; UBIQUITIN_2; 1.
 SQ SEQUENCE 245 AA; 26316 MW; 02C1D1B24C13139D CRC64;

Query Match 77.5%; Score 31; DB 11; Length 245;
 Best Local Similarity 77.8%; Pred. No. 1.2e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLPLQIL 9
 |||||
 Db 196 SLLPLLL 204

RESULT 22
 Q9BVT8
 ID Q9BVT8 PRELIMINARY; PRT; 246 AA.
 AC Q9BVT8;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)

DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE SIMILAR TO HYPOTHELICAL PROTEIN, CLONE 1-2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-PLACENTA, CHORIOCARCINOMA;
 RA Strausberg R.;
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC000936; AA000936.1; -;
 DR InterPro: IPR000626; Ubiquitin.
 DR SMART: SM00213; UBO; 1.
 DR PROSITE: PS00053; UBIQUITIN_2; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 246 AA; 26261 MW; E08E25A6B37665B3 CRC64;

 Query Match 77.5%; Score 31; DB 4; Length 246;
 Best Local Similarity 77.8%; Pred. No. 1.2e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 1 SLLPLQL 9
 Db 197 SLLPLQL 205

 RESULT 23
 076155 PRELIMINARY; PRT; 247 AA.
 AC 076155;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 26-KDA LECTIN (FRAGMENT).
 OS Periplaneta americana (American cockroach).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
 OC Blattodea; Blattellidae; Periplaneta.
 OX NCBI_TaxID=6978;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Arai T., Kawasaki K., Kubo T., Natori S.;
 RT "CDNA cloning of 26-kDa lectin."
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB012579; BA032800.1; -;
 DR HSP; P35247; I808.
 DR InterPro: IPR001304; lectin_c.
 DR Pfam: PF000059; lectin_c; 1.
 DR SMART: SM00034; CLECT; 1.
 DR PROSITE: PS00615; C_TYPE_LLECTIN_1; 1.
 DR PROSITE: PS50041; C_TYPE_LLECTIN_2; 1.
 KW Lectin.
 FT NON_TER
 SQ SEQUENCE 247 AA; 27575 MW; CFB4B488D03E5168 CRC64;

 Query Match 77.5%; Score 31; DB 5; Length 247;
 Best Local Similarity 77.8%; Pred. No. 1.2e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

 QY 1 SLLPLQL 9
 Db 19 SLLPLQL 27

 RESULT 24
 022910 PRELIMINARY; PRT; 303 AA.
 AC 022910;
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)

DE AT2G47360 PROTEIN.
 GN AT2G47360.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
 RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
 RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H., Moffat K.S.,
 RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
 RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
 RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
 RA Salzberg S.L., Fraser C.M., Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 thaliana."
 RL Nature 402:761-768 (1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC002337; AAB63836.1; -;
 SQ SEQUENCE 303 AA; 33965 MW; 12C5ADIA87573B4 CRC64;

 Query Match 77.5%; Score 31; DB 10; Length 303;
 Best Local Similarity 77.8%; Pred. No. 1.4e+02;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

 QY 1 SLLPLQL 9
 Db 29 SLLPLQL 37

 RESULT 25
 09V3Y4 PRELIMINARY; PRT; 316 AA.
 AC 09V3Y4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE MTCH PROTEIN (LD43650P).
 GN MTCH OR CG6851.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=107311132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Maniatis J., Venter L., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.I., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao O.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*,"
RL Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Jang J.S., Hahn Y., Park C., Chung J.H.;
RT "Identification of an evolutionary conserved mitochondrial carrier
RT family from various organisms,"
RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Y, CN BW SP;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA Gonzalez M., Guarin H., Li P., Lisao G., Miranda A., Mungall C.J.,
RA Nunoo J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AS003467; AAF47342.1; -;
DR EMBL: AF176011; AAD52649.1; -;
DR EMBL: AY051973; AAK93397.1; -;
DR FlyBase: FBgn0027786; Mtch.
DR InterPro: IPR001993; Mitoch_carrier.
DR Pfam: PF00153; mito_carr; 1.
SQ SEQUENCE 316 AA; 35011 MW; 4545D305A57A9991 CRC64;
Query Match 77.5%; Score 31; DB 5; Length 316;
Best Local Similarity 66.7%; Pred. NO. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLPLQL 9
DB 238 SLLPLQV 246
ID Q9MMS8 PRELIMINARY; PRT; 347 AA.
AC Q9MMS8
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CYTOCHROME B (FRAGMENT).
GN CYTB
OS Syngnathus sibilatrix.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Ciconiiformes; Ardeidae; Syrigma.
OX NCBI_TaxID=110687;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20188803; PubMed=10723744;
RA Sheldon F.H., Jones C.E., McCracken K.G.;
RT "Relative patterns and rates of evolution in heron nuclear and
RT mitochondrial DNA,"

RL Mol. Biol. Evol. 17:437-450(2000).
CC -!- FUNCTION: COMPONENT OF THE UBIQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
CC RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL
CC COUPLED TO ATP SYNTHESIS (BY SIMILARITY).
CC -!- COFACTOR: TWO HEME GROUPS (B562 AND B566) WHICH ARE NOT COVALENTLY
CC BOUND TO THE PROTEIN (BY SIMILARITY).
CC -!- SUBUNIT: THE MAIN SUBUNITS OF COMPLEX B-C1 ARE: CYTOCHROME B,
CC CYTOCHROME C1 AND THE RIESKE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CYTOCHROME B FAMILY.
CC EMBL: AF193820; AAF40124.1; -;
DR InterPro: IPR000179; Cyt_b_b6.
DR Pfam: PF00032; cytochrome_b_c; 1.
DR Pfam: PF00033; cytochrome_b_n; 1.
DR PROSITE: PS00192; CYTOCHROME_B_HEME; 1.
DR PROSITE: PS00193; CYTOCHROME_B_QO; UNKNOWN_1.
KW Electron transport; heme; Mitochondrion; Respiratory chain;
KW Transmembrane.
FT NON_TER 1
FT NON_TER 347 347
SQ SEQUENCE 347 AA; 38831 MW; 5FDC83D181366C51 CRC64;
Query Match 77.5%; Score 31; DB 8; Length 347;
Best Local Similarity 87.5%; Pred. NO. 1.6e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 LLLPLQL 9
DB 203 LLLPLML 210
ID Q9KALL PRELIMINARY; PRT; 392 AA.
AC Q9KALL
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BH2276 PROTEIN.
GN BH2276.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium *Bacillus*
RT *halodurans* and genomic sequence comparison with *Bacillus subtilis*,"
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001515; BAB05995.1; -;
KW Complete proteome.
SQ SEQUENCE 392 AA; 43310 MW; 12D0D32B6BE451DE CRC64;
Query Match 77.5%; Score 31; DB 16; Length 392;
Best Local Similarity 87.5%; Pred. NO. 1.8e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 LLLPLQL 9
DB 290 LLLPLQL 297
ID Q9KALL PRELIMINARY; PRT; 395 AA.
AC Q9KALL
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE YDER PROTEIN.
 GN YDER.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=1423;
 RN -[1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=168;
 RC MEDLINE=98000887; PubMed=9341680;
 RA BeIoan C., Ayora S., Exley R., Hirschbein L., Ogasawara N.,
 RA Kasahara Y., Alonso J.C., Le Hegarat F.;
 RT "Characterization of an lrp-like (lrpC) gene from Bacillus subtilis.";
 RL Mol. Gen. Genet. 256:63-71(1997).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=168;
 RC MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Mosser I., Albertini A.M., Alloni G.,
 RA Azevedo V., Berto M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borliss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guiseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaer-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Konigstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinols S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivoita C., Rocha E., Roche B., Rose M., Sadaie Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
 RT "The complete genome sequence of the gram-positive bacterium Bacillus
 subtilis.";
 RL Nature 390:249-256(1997).
 RN [3]
 RN SEQUENCE FROM N.A.
 RP STRAIN=168;
 RC MEDLINE=98000887; PubMed=9341680;
 RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB001488; BAA19365.1; -;
 DR EMBL; 299106; CAB12338.1; -;
 DR InterPro; IPR003662; sub.transporter.
 DR Pfam; PF00083; sugar_tr; 1.
 KW Complete proteome.
 SQ SEQUENCE 395 AA; 41144 MW; 9DEBBE05977D4CC8 CRC64;

Query Match 77.5%; Score 31; DB 16; Length 395;
 Best Local Similarity 87.5%; Pred. No. 1.9e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQI 8
 DB 294 SLLPLMI 301

RESULT 29
 Q98SC8

ID Q98SC8 PRELIMINARY; PRT; 401 AA.
 AC Q98SC8;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE SNF1-RELATED PROTEIN KINASE.
 GN KIN(SNF2).
 OS Guillardia theta (Cryptomonas phi).
 OG Nucleomorph.
 OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
 OX NCBI_TaxID=55529;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=21223349; PubMed=11323671;
 RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,
 RA Wu X., Reith M., Cavalier-Smith T., Maier U.G.;
 RT "The highly reduced genome of an enslaved algal nucleus.";
 RL Nature 410:1091-1096(2001).
 DR EMBL; AF083031; AAK39655.1; -;
 DR InterPro; IPR000719; Euk_pkinase.
 DR InterPro; IPR002290; Ser_thr_pkinase.
 DR Pfam; PF00069; pkinase; 1.
 DR PRINTS; PR00109; TYRKINASE.
 DR SMART; SM00220; S_TKC; 1.
 DR SMART; SM00219; TyRK; 1.
 DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
 KW ATP-binding; Kinase; Transferrase
 SQ SEQUENCE 401 AA; 45888 MW; 282CE9E6677CED4E CRC64;
 Query Match 77.5%; Score 31; DB 8; Length 401;
 Best Local Similarity 66.7%; Pred. No. 1.9e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 SLLPLQIL 9
 DB 304 SFLPLQLV 312
 RESULT 30
 O44312 PRELIMINARY; PRT; 438 AA.
 ID O44312
 AC O44312;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE GAG-LIKE ZINC-FINGER PROTEIN (FRAGMENT).
 GN R1-ELEMENT\ORF.
 OS Drosophila mercatorum mercatorum.
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=64960;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=99126433; PubMed=9927458;
 RA Malik H.S., Eickbush T.H.;
 RT "Retrotransposable elements R1 and R2 in the rDNA units of Drosophila
 mercatorum: abnormal abdomen revisited.";
 RL Genetics 151:653-665(1999).
 DR EMBL; AF015277; AAB94026.1; -;
 DR FlyBase; FBgn0044260; Dmer\R1-element\ORF.
 DR InterPro; IPR001878; znf_CCHC.
 DR Pfam; PF00098; znf_CCHC; 1.
 DR SMART; SM00343; Znf_C2HC; 2.
 KW Zinc; Zinc-finger.
 FT NON_TER 1
 SQ SEQUENCE 438 AA; 48470 MW; 05DFDE7BCA971F25 CRC64;
 Query Match 77.5%; Score 31; DB 5; Length 438;
 Best Local Similarity 77.8%; Pred. No. 2.1e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
||||| :|
Db 74 SLLPLLLL 82

RESULT 31
Q93I52 PRELIMINARY; PRT; 469 AA.
AC Q93I52;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE GLYCOSYLTRANSFERASE.
GN STHA.
OS Sphaerotilus natans.
OC Bacteria; Proteobacteria; beta subdivision; Comamonadaceae;
OC Sphaerotilus.
OX NCBI_TaxID=34103;
[1]
RN RN
RP RN
RC SRAIN=IAMI2068;
RA Suzuki T., Kanegawa T., Kamagata Y.;
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
[2]
RN RN
RP RN
RC SRAIN=IAMI2068;
RA Toshihiko S., Kanagawa T., Kamagata Y.;
RT "Identification of the Gene Essential for Sheathed Structure and Extracellular Polymer Formation in Sphaerotilus natans, a Filamentous RT Sheathed Bacterium."
RL Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB050638; BAB69050.1; -
KW Transferase.
SQ SEQUENCE 469 AA; 52271 MW; 73613E7EAD46C4D CRC64;

Query Match 77.5%; Score 31; DB 2; Length 469;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
||||| :|
Db 447 SLLDLQVL 455

RESULT 32
Q9SEA5 PRELIMINARY; PRT; 470 AA.
AC Q9SEA5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 57.8 KDA PROTEIN.
GN ORF470.
OS Guillardia theta (Cryptomonas phi).
OG Nucleomorph.
OC Eukaryota; Cryptophyta; Cryptomonadaceae; Guillardia.
OX NCBI_TaxID=55529;
[1]
RN RN
RP RN
RC SRAIN=IAMI2068;
RA Zauner S., Fraunholz M., Wastl J., Penny S., Beaton M.,
RA Cavalier-Smith T., Maier U.G., Douglas S.;
RT "Chloroplast protein and centrosomal genes, a tRNA intron, and odd telomeres in an unusually compact eukaryotic genome, the cryptomonad nucleomorph."
RL Proc. Natl. Acad. Sci. U.S.A. 97:200-205(2000).
[2]
RN RN
RP RN
RC SRAIN=IAMI2068;
RA Douglas S., Zauner S., Fraunholz M., Beaton M., Penny S., Deng L.T.,
RA Wu X., Reith M., Cavalier-Smith T., Maier U.G.;
RT "The highly reduced genome of an enslaved algal nucleus."
RL Nature 410:1091-1096(2001).

DR EMBL: AF165818; AAF24211.1; -
KW Hypothetical protein.
SQ SEQUENCE 470 AA; 57801 MW; 19FAA35E26A452FA CRC64;

Query Match 77.5%; Score 31; DB 8; Length 470;
Best Local Similarity 75.0%; Pred. No. 2.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
||||| :|
Db 253 MLLPLEIL 260

RESULT 33
O26479 PRELIMINARY; PRT; 475 AA.
AC O26479;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE O-ANTIGEN TRANSPORTER RELATED PROTEIN.
GN MTH379.
OS Methanothermobacter thermoautotrophicus.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
OC Methanothermobacter.
OX NCBI_TaxID=145262;
[1]
RN RN
RP RN
RC SRAIN=DELTA H;
RX Smith D.R., Doucette-Stamm L.A., DeLoughery C., Lee H.-M., DuBois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lumm W., Pothier B., Qiu D.,
RA Spadafora R., Vicare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiواني N., Caruso A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McDougall S., Shimer G., Goyal A., Pietrovski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT "Complete genome sequence of Methanobacterium thermoautotrophicum deltaH: functional analysis and comparative genomics."
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000823; AAB84885.1; -
DR InterPro: IPR002797; Polysacc_synt.
DR Pfam: PF01943; Polysacc_synt; 1.
DR PROSITE: PS00678; WD_REPEATS_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 475 AA; 51904 MW; 97852B34928608DD CRC64;

Query Match 77.5%; Score 31; DB 17; Length 475;
Best Local Similarity 87.5%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
||||| :|
Db 432 LLLPLSL 439

RESULT 34
Q9S335 PRELIMINARY; PRT; 519 AA.
AC Q9S335;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE DNA HELICASE (FRAGMENT).
GN UVRD.
OS Prochlorococcus marinus.
OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;
OC Prochlorococcus.
OX NCBI_TaxID=1219;
[1]
RN RN
RP RN
RC SRAIN=CCMP 1375;

RX MEDLINE=97008144; PubMed=8855320;
RA Hess W.R., Partensky F., van der Staay G.W., Garcia-Fernandez J.M.,
RA Boerner T., Vaulot D.;
RT "Coexistence of phycoerythrin and a chlorophyll a/b antenna in a
RT marine prokaryote.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:11126-11130(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CCMP 1375;
RX MEDLINE=99364545; PubMed=10437834;
RA Hess W.R., Steglich C., Lichtle C., Partensky F.;
RT "Phycoerythrins of the oxyphotobacterium Prochlorococcus marinus are
RT associated to the thylakoid membrane and are encoded by a single large
RT gene cluster.";
RL Plant Mol. Biol. 40:507-521(1999).
DR EMBL; AJ001230; CAB52698.1; -;
DR HSP; P56255; 1PJR
DR InterPro; IPR000212; UvrD-helicase.
DR Pfam; PF00580; UvrD-helicase; 1.
KW Helicase.
FT NON_TER 519 519
SQ SEQUENCE 519 AA; 59296 MW; 2F6F0B93A0DA57C6 CRC64;

Query Match 77.5%; Score 31; DB 2; Length 519;
Best Local Similarity 75.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
| | | | | : |
DB 231 LLLPVQL 238

RESULT 35
Q95Y88
ID Q95Y88 PRELIMINARY; PRT; 630 AA.
AC Q95Y88;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 69.0 KDA PROTEIN.
GN R13A5.9.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE=99069613; PubMed=9851916;
RA None;
RT "Genome sequence of the nematode C. elegans: a platform for
RT investigating biology. The C. elegans Sequencing Consortium.";
RL Science 282:2012-2018(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Macri C., Vaudin M.;
RT "The sequence of C. elegans cosmid R13A5.";
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RA Waterston R.;
RT "Direct Submission.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006679; AK84470.1; -;
KW Hypothetical protein.
SQ SEQUENCE 630 AA; 69009 MW; AC4696645631E750 CRC64;

Query Match 77.5%; Score 31; DB 5; Length 630;
Best Local Similarity 75.0%; Pred. No. 2.9e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
| | | | | : | | | | |
DB 473 MVLPQL 480

RESULT 36
Q9DER4
ID Q9DER4 PRELIMINARY; PRT; 934 AA.
AC Q9DER4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ZONA PELLUCIDA PROTEIN 1.
GN Zpl.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20435849; PubMed=10979984;
RA Bausek N., MacIawek M., Schneider W.J., Wohlrab F.;
RT "The major chicken egg envelope protein Zpl is different from ZPB and
RT is synthesized in the liver.";
RL J. Biol. Chem. 275:28866-28872(2000).
DR EMBL; AJ289697; CAC16087.1; -;
DR HSP; P04155; 1PS2.
DR InterPro; IPR000519; P_trefoil.
DR InterPro; IPR001507; zona_pellucida.
DR Pfam; PF00100; zona_pellucida; 1.
DR PRINTS; PR00023; ZPELLUCIDA.
DR SMART; SM00018; P: 1.
DR SMART; SM00241; ZP: 1.
SQ SEQUENCE 934 AA; 100028 MW; C28CB7743245B47F CRC64;

Query Match 77.5%; Score 31; DB 13; Length 934;
Best Local Similarity 77.8%; Pred. No. 4.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
| | | | | : |
DB 6 SLLPLLL 14

RESULT 37
Q910B7
ID Q910B7 PRELIMINARY; PRT; 934 AA.
AC Q910B7;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ZONA PELLUCIDA PROTEIN 1.
GN Zpl.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RA Pan J., Sasanami T., Suzuki Y., Mori M.;
RT "Effects of Estradiol on Production of Perivitelline Membrane Protein
RT Zpl in Liver of Japanese Quail.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB061520; BAB47585.1; -;
SQ SEQUENCE 934 AA; 100508 MW; C3F51AFB4190064C CRC64;

Query Match 77.5%; Score 31; DB 13; Length 934;
Best Local Similarity 77.8%; Pred. No. 4.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

||||| :1
6 SLLPLLLL 14

Db 6 SLLPLLLL 14
Query Match 75.0%; Score 30; DB 15; Length 82;
Best Local Similarity 85.7%; Pred. No. 66;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 38

Q9GJR4 ID Q9GJR4 PRELIMINARY; PRT; 69 AA.
AC Q9GJR4;
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE INTERLEUKIN 2 PRECURSOR (FRAGMENT).
GN IL-2.
OS Ovis aries (Sheep).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SHEEP 2, AND SHEEP 1;
RA Luehken G., Prinzenberg E.-M., Hiendleder S., Erhardt G.;
RT "A single strand conformation polymorphism in the ovine interleukin 2
(IL-2) gene.";
RL J. Anim. Sci. 0:0-0(2000).
DR EMBL; AF215687; AAC43986.1; -;
DR EMBL; AF213883; AAC35709.1; -;
DR HSP; P01585; IRL
DR InterPro: IPR000779; Interleukin-2.
DR PRINTS; PR00265; INTERLEUKIN2.
DR ProDom: PD003649; Interleukin-2; 1.
DR SMART; SM00189; IL2; 1.
KW Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 >69 INTERLEUKIN 2.
FT NON_TER 69 69
SQ SEQUENCE 69 AA; 7711 MW; B8768C23BE34D1AE CRC64;

Query Match 75.0%; Score 30; DB 6; Length 69;
Best Local Similarity 77.8%; Pred. No. 56;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
||||| :1
Db 36 SLLPLQLL 44

RESULT 39

Q9WI77 ID Q9WI77 PRELIMINARY; PRT; 82 AA.
AC Q9WI77;
DT 01-NOV-1999 (TReMBLrel. 12, Created)
DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE VPU PROTEIN.
GN VPU.
OS Human immunodeficiency virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TWAL;
RX MEDLINE=20341532; PubMed=10878027;
RA Lee C.N., Wang W.K., Fan W.S., Twu S.J., Chen S.C., Sheng M.C.,
Chen M.Y.;
RT "Determination of Human Immunodeficiency Virus Type 1 Subtypes in
Taiwan by vpu Gene Analysis."
RL J. Clin. Microbiol. 38:2468-2474(2000).
DR EMBL; AF143901; AAD34575.1; -;
DR HSP; P19554; VPU.
DR InterPro: IPR002094; Vpu.
DR Pfam: PF00558; Vpu; 1.
SQ SEQUENCE 82 AA; 9255 MW; 5A9E57FFAFA64251 CRC64;

Qy 2 LLLPLQL 8
:|||||
Db 1 LLLPLQL 7

RESULT 40

Q9DOC8 ID Q9DOC8 PRELIMINARY; PRT; 121 AA.
AC Q9DOC8;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE 2610027F03RIK PROTEIN.
GN 2610027F03RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
Hayashizaki Y.
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK011565; BAB27702.1; -;
DR MGD; MGI:1917154; 2610027F03RIK.
SQ SEQUENCE 121 AA; 13948 MW; F4BAB890F461D6C CRC64;

Query Match 75.0%; Score 30; DB 11; Length 121;
Best Local Similarity 75.0%; Pred. No. 96;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
:||||| :1
Db 96 LLLPLRL 103

RESULT 41

Q9CKE2 ID Q9CKE2 PRELIMINARY; PRT; 150 AA.
AC Q9CKE2;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN PM1680.
GN PM1680.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;

```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida pm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006204; AAK03764.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 150 AA; 17314 MW; 7454710498463FF9 CRC64;

Query Match 75.0%; Score 30; DB 16; Length 150;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
Db 24 LLVPLQIV 31

RESULT 42
Q9GL83
ID Q9GL83 PRELIMINARY; PRT; 155 AA.
AC Q9GL83;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTERLEUKIN 2.
GN IL-2.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE FROM N.A.
RA Ying Q.H., Li X.R., Pan J.Y.;
RT "Cloning of the goat IL-2 gene and its expression in E.coli.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE307018; AAG28783.1; -.
DR HSSP; P01585; 3INK.
DR InterPro; IPR000779; Interleukin-2.
DR Pfam; PF00715; IL2; 1.
DR PRINTS; PR00265; INTERLEUKIN2.
DR ProDom; PD003649; Interleukin-2; 1.
DR SMART; SM00189; IL2; 1.
DR PROSITE; PS00424; INTERLEUKIN_2; 1.
SQ SEQUENCE 155 AA; 17605 MW; EEB2DE18F5469AA CRC64;

Query Match 75.0%; Score 30; DB 6; Length 155;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 36 SLLLDLQL 44

RESULT 43
Q95MP4
ID Q95MP4 PRELIMINARY; PRT; 155 AA.
AC Q95MP4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTERLEUKIN 2 PRECURSOR.
GN IL-2.
OS Ovis aries (Sheep).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Ovis.
OX NCBI_TaxID=9940;
RN [1]
RP SEQUENCE FROM N.A.
RA Luehken G., Prinzenberg E., Hiendler S., Erhardt G.;
RT "A single strand conformation polymorphism in the ovine Interleukin 2
(IL-2) gene.";
RL J. Anim. Sci. 0:0-0(2000).
DR EMBL; AF287479; AAK69554.1; -.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 22 155 INTERLEUKIN 2.
FT CHAIN 1 21 POTENTIAL.
SQ SEQUENCE 155 AA; 17678 MW; 3F76C6CBB7B289C5 CRC64;

Query Match 75.0%; Score 30; DB 6; Length 155;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 36 SLLLDLQL 44

RESULT 44
Q95KP3
ID Q95KP3 PRELIMINARY; PRT; 155 AA.
AC Q95KP3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTERLEUKIN-2.
OS Bubalus bubalis (Domestic water buffalo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bubalus.
OX NCBI_TaxID=89462;
RN [1]
RP SEQUENCE FROM N.A.
RA Sreekumar E., Premraj A., Saravanakumar M., Rasool T.J.;
RT "Full length cDNA of buffalo interleukin-2.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF363786; AAK50039.1; -.
SQ SEQUENCE 155 AA; 17541 MW; 1753173FE619946B CRC64;

Query Match 75.0%; Score 30; DB 6; Length 155;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 36 SLLLDLQL 44

RESULT 45
Q9IT40
ID Q9IT40 PRELIMINARY; PRT; 162 AA.
AC Q9IT40;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 18.8 KDA PROTEIN.
OS lumpy skin disease virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Capripoxvirus.
OX NCBI_TaxID=59509;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEETHLING;
RA Stipinovich C., Vreede F.T., Kara P.D., Wallace D.B., Nel L.H.,
RA Viljoen G.J.;
RT "Molecular characterization of important regions of the lumpy skin
disease virus genome.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF336128; AAK43550.1; -.
KW Hypothetical protein.
SQ SEQUENCE 162 AA; 18782 MW; 5F914A4080F729EE CRC64;
```

```

RP SEQUENCE FROM N.A.
RA Luehken G., Prinzenberg E., Hiendler S., Erhardt G.;
RT "A single strand conformation polymorphism in the ovine Interleukin 2
(IL-2) gene.";
RL J. Anim. Sci. 0:0-0(2000).
DR EMBL; AF287479; AAK69554.1; -.
KW SIGNAL.
FT SIGNAL.
FT CHAIN 22 155 INTERLEUKIN 2.
FT CHAIN 1 21 POTENTIAL.
SQ SEQUENCE 155 AA; 17678 MW; 3F76C6CBB7B289C5 CRC64;

Query Match 75.0%; Score 30; DB 6; Length 155;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 36 SLLLDLQL 44

RESULT 44
Q95KP3
ID Q95KP3 PRELIMINARY; PRT; 155 AA.
AC Q95KP3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INTERLEUKIN-2.
OS Bubalus bubalis (Domestic water buffalo).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bubalus.
OX NCBI_TaxID=89462;
RN [1]
RP SEQUENCE FROM N.A.
RA Sreekumar E., Premraj A., Saravanakumar M., Rasool T.J.;
RT "Full length cDNA of buffalo interleukin-2.";
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF363786; AAK50039.1; -.
SQ SEQUENCE 155 AA; 17541 MW; 1753173FE619946B CRC64;

Query Match 75.0%; Score 30; DB 6; Length 155;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 36 SLLLDLQL 44

RESULT 45
Q9IT40
ID Q9IT40 PRELIMINARY; PRT; 162 AA.
AC Q9IT40;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 18.8 KDA PROTEIN.
OS lumpy skin disease virus.
OC Viruses; dsDNA viruses, no RNA stage; Poxviridae; Chordopoxvirinae;
OC Capripoxvirus.
OX NCBI_TaxID=59509;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEETHLING;
RA Stipinovich C., Vreede F.T., Kara P.D., Wallace D.B., Nel L.H.,
RA Viljoen G.J.;
RT "Molecular characterization of important regions of the lumpy skin
disease virus genome.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF336128; AAK43550.1; -.
KW Hypothetical protein.
SQ SEQUENCE 162 AA; 18782 MW; 5F914A4080F729EE CRC64;
```

Query Match 75.0%; Score 30; DB 12; Length 162;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
| | | | |
Db 90 LLLPLTIL 97

Search completed: November 6, 2002, 12:12:09
Job time : 20.4444 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 5.33333 Seconds
(without alignments)
65.339 Million cell updates/sec

Title: US-09-905-083-36
Perfect score: 40
Sequence: 1 SLLPLQL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	40	100.0	233	1	KLK7_HUMAN
2	34	85.0	361	1	IHA_TRIUVU
3	32	80.0	494	1	EXOT_RHIME
4	31	77.5	202	1	CUTI_BOTCI
5	31	77.5	354	1	GPDA_TRYBB
6	31	77.5	354	1	GPDA_TRYBR
7	31	77.5	428	1	YFJD_ECOLI
8	31	77.5	783	1	KMHC_DICDI
9	30	75.0	155	1	IL2_BOVIN
10	30	75.0	155	1	IL2_CAPHI
11	30	75.0	155	1	IL2_SHEEP
12	30	75.0	162	1	IL2_CEREL
13	30	75.0	289	1	CTR4_SCHPO
14	30	75.0	333	1	CD1B_HUMAN
15	30	75.0	369	1	PSPD_BOVIN
16	30	75.0	371	1	CONG_BOVIN
17	30	75.0	374	1	YHJ1_ECOLI
18	30	75.0	435	1	UL49_HSVSA
19	30	75.0	463	1	YICE_ECOLI
20	30	75.0	524	1	K2C4_MOUSE
21	30	75.0	524	1	PBBT_MOUSE
22	30	75.0	573	1	COX1_MYCTU
23	30	75.0	592	1	NDC1_HUMAN
24	30	75.0	783	1	RINT_HUMAN
25	29	72.5	126	1	NRM1_RAT
26	29	72.5	198	1	CD8A_PONPY
27	29	72.5	208	1	NU4M_PHOSU
28	29	72.5	235	1	CD8A_HUMAN
29	29	72.5	239	1	CD8A_CANFA
30	29	72.5	239	1	FLGH_AGRF5
31	29	72.5	242	1	ATPI_ODOSI
32	29	72.5	243	1	CD48_HUMAN
33	29	72.5	306	1	TRUB_NEIMA

34	29	72.5	307	1	TRUB_NEIMB
35	29	72.5	397	1	O59A_DROME
36	29	72.5	434	1	SECY_BORBU
37	29	72.5	435	1	YA06_HUMAN
38	29	72.5	462	1	PTGR_PEARF
39	29	72.5	477	1	PTGB_BUCAI
40	29	72.5	477	1	PTGB_ECOLI
41	29	72.5	477	1	PTGB_SALTY
42	29	72.5	500	1	YA04_HUMAN
43	29	72.5	538	1	NRM1_PIG
44	29	72.5	538	1	YA94_HUMAN
45	29	72.5	548	1	NRM1_BISBI

ALIGNMENTS

RESULT 1
KLK7_HUMAN STANDARD; PRT; 253 AA.
AC P49862:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic enzyme) (hSCCE).
DE DE
GN KLK7 OR PRSS6 OR SCCE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.
RC TISSUE=Skin.
RX MEDLINE=94308225; PubMed=8034709;
RA Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A., Egelrud T.;
RT "Cloning, expression, and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase.";
RL J. Biol. Chem. 269:19420-19426(1994).
[2]
RN SEQUENCE FROM N.A.
RP TISSUE=Keratinocytes;
RA Yousef G.M., Scorillas A., Diamandis E.P.;
RT "Molecular characterization, mapping and tissue expression of the human stratum corneum chymotryptic enzyme gene.";
RL Submitted (DEC 1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RX MEDLINE=20510030; PubMed=11054574;
RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J., Moss P., Paepel B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene cluster located in chromosome 19q13 region.";
RL Gene 257:119-130(2000).
[4]
RN CHARACTERIZATION.
RX MEDLINE=95314630; PubMed=7794273;
RA Skjott A., Stroemqvist M., Egelrud T.;
RT "Primary substrate specificity of recombinant human stratum corneum chymotryptic enzyme.";
RL Biochem. Biophys. Res. Commun. 211:586-589(1995).
CC -!- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE CLEAVES INSULIN B CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE-1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION OF PRECURSORS TO INFLAMMATORY CYTOKINES.
CC -!- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE ALSO SEEN IN THE BRAIN AND KIDNEY.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

```
CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L33404; AAC37551.1; -
CC EMBL: AF166330; AAD49718.1; -
CC EMBL: AF243527; AAG33360.1; -
CC HSSP: P00763; IDPO.
CC MEROPS: S01.300; -
CC MW: 604438; -
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00020; Tryp_SPC; 1.
CC PROSITE: PS0240; TRYPsin_DOM; 1.
CC PROSITE: PS00134; TRYPsin_HIS; 1.
CC PROSITE: PS00135; TRYPsin_SER; 1.
CC Hydrolase: Serine protease; Zymogen; Glycoprotein; Signal.
CC -----
FT SIGNAL 1 22
FT PROPEP 23 29 ACTIVATION PEPTIDE.
FT CHAIN 30 253 KALLIKREIN 7.
FT ACT_SITE 70 70 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 36 137 BY SIMILARITY.
FT DISULFID 55 71 BY SIMILARITY.
FT DISULFID 137 239 BY SIMILARITY.
FT DISULFID 144 211 BY SIMILARITY.
FT DISULFID 176 190 BY SIMILARITY.
FT DISULFID 201 226 BY SIMILARITY.
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 253 AA; 27525 MW; 2D68B6B15A76A668 CRC64;

Query Match 100.0%; Score 40; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.62;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
Db 4 SLLPLQL 12

RESULT 2
ID IHA_TRIVU STANDARD; PRT; 361 AA.
AC O77755;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Inhibin alpha chain precursor.
GN INHA.
OS Trichosurus vulpecula (Brush-tailed possum).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Metatheria; Diprotodontia; Phalangeridae; Trichosurus.
OX NCBI_TaxID=9337;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99027340; PubMed=9801457;
RA Vannontfort D., Fidler A.E., Heath D.A., Lawrence S.B., Tisdall D.J.,
RA Greenwood P.J., McNatty K.;
RA "cDNA sequence analysis, gene expression and protein localisation of
RA the inhibin alpha subunit of Australian brushtail possum (Trichosurus
RA vulpecula).";
RL J. Mol. Endocrinol. 21:141-152(1998).
CC -!- FUNCTION: INHIBINS AND ACTIVINS INHIBIT AND ACTIVATE,
CC RESPECTIVELY, THE SECRETION OF FOLLITROPIN BY THE PITUITARY GLAND.
```

```
CC INHIBINS/ACTIVINS ARE INVOLVED IN REGULATING A NUMBER OF DIVERSE
CC FUNCTIONS SUCH AS HYPOTHALAMIC AND PITUITARY HORMONE SECRETION,
CC GONADAL HORMONE SECRETION, GERM CELL DEVELOPMENT AND MATURATION,
CC ERYTHROID DIFFERENTIATION, INSULIN SECRETION, NERVE CELL SURVIVAL,
CC EMBRYONIC AXIAL DEVELOPMENT OR BONE GROWTH, DEPENDING ON THEIR
CC SUBUNIT COMPOSITION. INHIBINS APPEAR TO OPPOSE THE FUNCTIONS OF
CC ACTIVINS.
CC -!- SUBUNIT: DIMERIC, LINKED BY ONE OR MORE DISULFIDE BONDS.
CC INHIBIN A IS A DIMER OF ALPHA AND BETA-A.
CC INHIBIN B IS A DIMER OF ALPHA AND BETA-B.
CC -!- SIMILARITY: BELONGS TO THE TGF-BETA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF033340; AAC63945.1; -
CC InterPro: IPR002405; Inhibin_alpha.
CC InterPro: IPR001839; TGF-beta.
CC Pfam: PF00019; TGF-beta; 1.
CC PRINTS: PR00669; INHIBINA.
CC ProDom: PD000357; TGF-beta; 1.
CC SMART: SM00204; TGF-beta; 1.
CC PROSITE: PS00250; TGF-BETA_1; 1.
CC Growth factor; Hormone; Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT PROPEP 22 230 BY SIMILARITY.
FT CHAIN 231 361 INHIBIN ALPHA CHAIN.
FT DISULFID 260 323 BY SIMILARITY.
FT DISULFID 289 358 BY SIMILARITY.
FT DISULFID 293 360 BY SIMILARITY.
FT DISULFID 322 322 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 48 48 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 266 266 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 361 AA; 38945 MW; D661C0F93CDA87D CRC64;

Query Match 85.0%; Score 34; DB 1; Length 361;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
Db 4 LLLPLQL 11

RESULT 3
ID EXOT_RHIME STANDARD; PRT; 494 AA.
AC P33699;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Succinoglycan biosynthesis transport protein exot.
GN EXOT OR RB1074 OR SMB20950.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OX Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94042869; PubMed=8226645;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RA "Family of glycosyl transferases needed for the synthesis of
RA succinoglycan by Rhizobium meliloti.";
RL J. Bacteriol. 175:7033-7044(1993).
RN [2]
```

RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=94042870; PubMed=8226646;
 RA Gluckmann M.A., Reuber T.L., Walker G.C.;
 RT "Genes needed for the modification, polymerization, export, and
 processing of succinoglycan by Rhizobium meliloti: a model for
 succinoglycan biosynthesis.";
 RL J. Bacteriol. 175:7045-7055(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RCR2011 / SU47;
 RX MEDLINE=94162682; PubMed=8118055;
 RA Becker A., Kleickmann A., Kuester H., Keller M., Arnold W.,
 RA Puehler A.;
 RT "Analysis of the Rhizobium meliloti genes exoU, exoV, exoW, exoT, and
 exoI involved in exopolysaccharide biosynthesis and nodule invasion:
 exoU and exoV probably encode glucosyltransferases.";
 RL Mol. Plant Microbe Interact. 6:735-744(1993).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1021;
 RX MEDLINE=21396508; PubMed=11481431;
 RA Finan T.M., Weidner S., Mong K., Buhrmester J., Chain P.,
 RA Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
 RA Golding B., Puehler A.;
 RT "The complete sequence of the 1.683-kb pSymB megaplasmid from the N2-
 fixing endosymbiont Sinorhizobium meliloti.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
 RC -1- PATHWAY: EXOPOLYSACCHARIDE BIOSYNTHESIS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
 CC -1- SIMILARITY: BELONGS TO THE POLYSACCHARIDE SYNTHASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: L20758; AAA16050.1; -;
 DR EMBL: 222646; CAA80362.1; -;
 DR EMBL: AL603645; CAC49474.1; -;
 DR PIR: C49349; C49349.
 DR InterPro: IPR002797; Polysacc synt.
 DR Pfam: PF01943; Polysacc synt. 1
 DR Exopolysaccharide synthesis; Transmembrane; Plasmid;
 KW Complete proteome.
 FT TRANSMEM 16 36 POTENTIAL.
 FT TRANSMEM 44 64 POTENTIAL.
 FT TRANSMEM 82 102 POTENTIAL.
 FT TRANSMEM 105 125 POTENTIAL.
 FT TRANSMEM 157 177 POTENTIAL.
 FT TRANSMEM 215 235 POTENTIAL.
 FT TRANSMEM 253 273 POTENTIAL.
 FT TRANSMEM 297 317 POTENTIAL.
 FT TRANSMEM 321 341 POTENTIAL.
 FT TRANSMEM 343 363 POTENTIAL.
 FT TRANSMEM 384 404 POTENTIAL.
 FT TRANSMEM 421 441 POTENTIAL.
 FT TRANSMEM 447 467 POTENTIAL.
 SQ SEQUENCE 494 AA; 53315 MW; 008521D05FDB6013 CRC64;

Query Match 80.08; Score 32; DB 1; Length 494;
 Best Local Similarity 75.08; Pred. No. 48;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQI 8
 |||:|:
 Db 122 SLLPLQL 129

RESULT 4

CUTI_BOTCI STANDARD; PRT; 202 AA.
 ID CUTI_BOTCI
 AC Q00298;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Cutinase precursor (EC 3.1.1.74).
 GN CUTA.
 OS Botrytis cinerea (Botryotinia fuckeliana).
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Leotiomycetes;
 OC Helotiales; Sclerotiniaceae; Botryotinia.
 OX NCBI_TaxID=40559;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAS56;
 RX MEDLINE=97155570; PubMed=9002269;
 RA van der Vlugt-Bergmans C.J.B., Wagmakers L.C.A.M., van Kan J.A.L.;
 RT "Cloning and expression of the cutinase A gene of Botrytis cinerea.";
 RL Mol. Plant Microbe Interact. 10:21-29(1997).
 CC -1- FUNCTION: CATALYZE THE HYDROLYSIS OF CUTIN, A POLYESTER THAT FORMS
 THE STRUCTURE OF PLANT CUTICLE. ALLOWS PATHOGENIC FUNGI TO
 PENETRATE THROUGH THE CUTICULAR BARRIER INTO THE HOST PLANT DURING
 THE INITIAL STAGE OF THE FUNGAL INFECTION (By similarity).
 CC -1- CATALYTIC ACTIVITY: Cutin + H(2)O = cutin monomers.
 CC -1- SUBCELLULAR LOCATION: Secreted (By similarity).
 CC -1- PTM: THE 2 DISULFIDE BONDS PLAY A CRITICAL ROLE IN HOLDING THE
 CATALYTIC RESIDUES IN JUXTA-POSITION: REDUCTION OF THE DISULFIDE
 BRIDGES RESULTS IN THE COMPLETE INACTIVATION OF THE ENZYME (BY
 SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CUTINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: Z69264; CAA93255.1; -;
 DR HSSP; P00590; IXZJ.
 DR InterPro: IPR000675; Cutinase.
 DR Pfam: PF01083; Cutinase; 1.
 DR PROSITE; PS00155; CUTINASE_1; 1.
 DR PROSITE; PS00931; CUTINASE_2; 1.
 DR Hydrolase; Serine esterase; Signal.
 KW SIGNAL 1 20 POTENTIAL.
 FT CHAIN 21 202 CUTINASE.
 FT DISULFID 31 172 BY SIMILARITY.
 FT DISULFID 106 165 BY SIMILARITY.
 FT ACT_SITE 117 117 BY SIMILARITY.
 FT ACT_SITE 169 169 BY SIMILARITY.
 FT ACT_SITE 182 182 BY SIMILARITY.
 SQ SEQUENCE 202 AA; 20253 MW; 031D64725A23D8E7 CRC64;
 Query Match 77.5%; Score 31; DB 1; Length 202;
 Best Local Similarity 66.7%; Pred. No. 31;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 SLLPLQL 9
 :|:|:|:|:
 Db 11 ALLPLSVL 19
 RESULT 5
 GPDA_TRYBB STANDARD; PRT; 354 AA.
 ID GPDA_TRYBB
 AC P90593;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE GlyceroL-3-phosphate dehydrogenase [NAD+], glycosomal (EC 1.1.1.8).
 GN GPD.

```

OS Trypanosoma brucei brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OC NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=427;
RX MEDLINE=97077437; PubMed=8920004;
RA Kohl L., Dmota T., Thi C.D., Callens M., van Beeumen J.,
RA Opperdoes F.R., Michels P.A.M.;
RT "Cloning and characterization of the NAD-linked glycerol-3-phosphate
RT dehydrogenases of Trypanosoma brucei brucei and Leishmania mexicana
RT mexicana and expression of the trypanosome enzyme in Escherichia
RT coli.";
RL Mol. Biochem. Parasitol. 76:159-173(1996).
CC -|- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(+) = glycerone
CC phosphate + NADH.
CC -|- SUBCELLULAR LOCATION: Glycosomal.
CC -|- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
CC DEHYDROGENASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X89738; CAA61890.1; -.
DR InterPro; IPR001652; NAD_Gly3P_dh.
DR Pfam; PF01210; NAD_Gly3P_dh; 1.
DR PRINTS; PR00077; GPDHGRGNASE.
DR PRODOM; PD001649; NAD_Gly3P_dh; 1.
DR PROSITE; PS00957; NAD_G3PDH; FALSE_NEG.
KW Oxidoreductase; NAD; Glycosome.
FT SITE 352 354 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 354 AA; 37783 MW; 12CE19AEDA9E4EC9 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 354;
Best Local Similarity 77.8%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
DB 124 SLLFPAQIL 132

RESULT 6
GPDA_TRYBR STANDARD; PRT; 354 AA.
AC Q26756;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Glycerol-3-phosphate dehydrogenase [NAD+], glycosomal (EC 1.1.1.8).
GN GPD.
OS Trypanosoma brucei rhodesiense.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=31286;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VITAT 1.1;
RX MEDLINE=97077436; PubMed=8920003;
RA Stebeck C.E., Frevert U., Mommensen T.P., Vassella E., Roditi I.,
RA Pearson T.W.;
RT "Molecular characterization of glycosomal NAD(+)-dependent glycerol 3-
RT phosphate dehydrogenase from Trypanosoma brucei rhodesiense.";
RL Mol. Biochem. Parasitol. 76:145-158(1996).
CC -|- CATALYTIC ACTIVITY: Sn-glycerol 3-phosphate + NAD(+) = glycerone
CC phosphate + NADH.
CC -|- SUBCELLULAR LOCATION: Glycosomal.
CC -|- SIMILARITY: BELONGS TO THE NAD-DEPENDENT GLYCEROL-3-PHOSPHATE
CC DEHYDROGENASE FAMILY.

```

```

CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X91142; CAA62581.1; -.
DR InterPro; IPR001652; NAD_Gly3P_dh.
DR Pfam; PF01210; NAD_Gly3P_dh; 1.
DR PRINTS; PR00077; GPDHGRGNASE.
DR PRODOM; PD001649; NAD_Gly3P_dh; 1.
DR PROSITE; PS00957; NAD_G3PDH; FALSE_NEG.
KW Oxidoreductase; NAD; Glycosome.
FT SITE 352 354 MICROBODY TARGETING SIGNAL (POTENTIAL).
SQ SEQUENCE 354 AA; 37834 MW; 0CE049C5E8F83C00 CRC64;

Query Match 77.5%; Score 31; DB 1; Length 354;
Best Local Similarity 77.8%; Pred. No. 55;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
DB 124 SLLFPAQIL 132

RESULT 7
YFJD_ECOLI STANDARD; PRT; 428 AA.
AC P37908; P76600; P76601; P77009;
DT 01-OCT-1994 (Rel. 30, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yfjd.
GN YFJD OR B2612/B2613.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.P.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sempel G., Satoh Y., Sivasubram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
RN [3]
RP SEQUENCE OF 322-428 FROM N.A.
RC STRAIN=B178;
RX MEDLINE=88319942; PubMed=3045760;
RA Lipinska B., King J., Ang B., Georgopoulos C.;
RT "Sequence analysis and transcriptional regulation of the Escherichia
RT coli grpE gene, encoding a heat shock protein.";
RL Nucleic Acids Res. 16:7545-7562(1988).

```



```

RN [4]
RP Ridd K.E.;
RL Unpublished observations (AUG-1994).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE UPF0053 FAMILY. STRONG, TO
CC H. INFLUENZAE HI1017.
CC -1- SIMILARITY: CONTAINS 2 CBS DOMAINS.
CC -1- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC FRAMESHIFT IN POSITION 199 THAT PRODUCES TWO ORFS.
CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO NUMEROUS
CC FRAMESHIFTS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000347; AAC75661.1; ALT_FRAME.
CC EMBL; AE000347; AAC75662.1; ALT_FRAME.
CC EMBL; D90888; BAA16497.1; ALT_INIT.
CC EMBL; X07863; -; NOT_ANNOTATED_CDS.
CC EcoGene; EG12442; yfjD.
CC InterPro; IPR000644; CBS.
CC InterPro; IPR002550; DUF21.
CC Pfam; PF00571; CBS; 2.
CC Pfam; PF01595; DUF21; 1.
CC SMART; SM00116; CBS; 1.
CC KW Hypothetical protein; CBS domain; Repeat; Transmembrane;
CC Complete proteome. 24
CC TRANSNEM 4 POTENTIAL.
CC TRANSNEM 65 POTENTIAL.
CC TRANSNEM 92 112
CC TRANSNEM 130 150 POTENTIAL.
CC SEQUENCE 428 AA; 48044 MW; 38456865EDBCB151 CRC64;
Query Match 77.5%; Score 31; DB 1; Length 428;
Best Local Similarity 87.5%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 2 LLLPLQLL 9
DB 131 LLAPLQLL 138
ID KMHCDI STANDARD; PRT; 783 AA.
AC P34125;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain kinase (EC 2.7.1.129) (MHCK).
GN MHCK.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
ON NCBI_TaxID=44689;
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=92335200; PubMed=1321427;
RA David S., Spudich J.A.;
RT "Membrane-bound Dictyostelium myosin heavy chain kinase: a
RT developmentally regulated substrate-specific member of the protein
RT kinase C family.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:5877-5881(1992).
CC -1- FUNCTION: BRINGS ABOUT PHOSPHORYLATION OF SPECIFIC THREONINES IN
CC THE HEAVY CHAINS OF MYOSIN DURING CHEMOTAXIS, WHICH RESULTS IN
CC INHIBITION OF MYOSIN THICK FILAMENT FORMATION. PHOSPHORYLATION
CC ALSO INHIBITS THE ACTIN-ACTIVATED ATPASE ACTIVITY OF MYOSIN.

```

```

CC -1- CATALYTIC ACTIVITY: ATP + [myosin heavy-chain] - ADP + [myosin
CC heavy-chain] phosphate.
CC -1- SUBCELLULAR LOCATION: Membrane-bound.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ONLY DURING DEVELOPMENT.
CC -1- SIMILARITY: BELONGS TO THE EUKARYOTIC DIACYLGLYCEROL KINASE
CC FAMILY.
CC -1- SIMILARITY: CONTAINS 2 ZINC-DEPENDENT PHORBOL-ESTER AND DAG
CC BINDING DOMAINS.
CC -1- CAUTION: WHILE THIS IS EXPERIMENTALLY DEIGNED TO BE A PROTEIN
CC KINASE, IT COULD BE, ON THE BASIS OF SEQUENCE SIMILARITIES, A
CC DIACYLGLYCEROL KINASE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M93393; -; NOT_ANNOTATED_CDS.
CC PIR; A46136; A46136.
CC DictyDb; DD01046; mhck.
CC InterPro; IPR000756; DAGKa.
CC InterPro; IPR001206; DAGKc.
CC InterPro; IPR002219; DAG_PE-bind.
CC Pfam; PF00609; DAGKa; 1.
CC Pfam; PF00781; DAGKc; 1.
CC ProDom; PD002939; DAGKa; 1.
CC SMART; SM00109; Cl. 2.
CC SMART; SM00045; DAGKa; 1.
CC SMART; SM00046; DAGKc; 1.
CC PROSITE; PS00479; DAG_PE_BIND_DOM_1; 1.
CC PROSITE; PS00081; DAG_PE_BIND_DOM_2; 2.
CC Transferrase; Serine/threonine-protein kinase; ATP-binding; Membrane;
CC Phosphorylation; Phorbol-ester binding; Repeat.
CC DOMAIN 19 69 PHORBOL-ESTER AND DAG BINDING 1.
CC DOMAIN 88 138 PHORBOL-ESTER AND DAG BINDING 2.
CC NP_BIND 466 474 ATP (BY SIMILARITY).
CC BINDING 489 489 ATP (BY SIMILARITY).
CC DOMAIN 188 196 ASN-RICH.
CC DOMAIN 754 773 SER/THR-RICH.
CC MOD_RES 768 768 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
CC SEQUENCE 783 AA; 89154 MW; 6CAF177F4436F43E CRC64;
Query Match 77.5%; Score 31; DB 1; Length 783;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLPLQLL 9
DB 236 SLLNPLQLL 244
ID IL2_BOVIN STANDARD; PRT; 155 AA.
AC P05016;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
GN IL2 OR IL-2.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
ON NCBI_TaxID=9913;
RP SEQUENCE FROM N.A.
RC STRAIN=
RX MEDLINE=86205869; PubMed=3517854;
RA Carretti D.P., McKereghan K., Larsen A., Cantrell M.A., Anderson D.,

```

```
RA Gillis S., Cosman D., Baker P.E.;
RT "Cloning, sequence, and expression of bovine interleukin 2.";
RN Proc. Natl. Acad. Sci. U.S.A. 83:3223-3227(1986).
RX [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-86205870; PubMed-3486415;
RA Reeves R., Spies A.G., Nissen M.S., Buck C.D., Weinberg A.D.,
RA Barr P.J., Magnuson N.S., Magnuson J.A.;
RT "Molecular cloning of a functional bovine interleukin 2 cDNA.";
RN Proc. Natl. Acad. Sci. U.S.A. 83:3228-3232(1986).
RX [3]
RP SEQUENCE OF 1-22 FROM N.A.
RC TISSUE=Thymus;
RA Anikeeva N.N., Vinogradova T.V., Votoshin O.N.;
RA Submitted (DEC-1989) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PRODUCED BY T-CELLS IN RESPONSE TO ANTIGENIC OR
CC MITOGENIC STIMULATION. THIS PROTEIN IS REQUIRED FOR T-CELL
CC PROLIFERATION AND OTHER ACTIVITIES CRUCIAL TO REGULATION OF THE
CC IMMUNE RESPONSE. CAN STIMULATE B CELLS, MONOCYTES, LYMPHOKINE-
CC ACTIVATED KILLER CELLS, NATURAL KILLER CELLS, AND GLIOMA CELLS.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: BELONGS TO THE IL-2 FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12791; AAA30586.1; -
DR EMBL; M13204; AAA21143.1; ALT_INIT.
DR EMBL; X17201; CAA35062.1; -
DR EMBL; X52687; CAA36912.1; -
DR HSP; P01585; 3INK.
DR InterPro; IPR000779; Interleukin-2.
DR Pfam; PF00715; IL2; 1.
DR PRINTS; PR00265; INTERLEUKIN2.
DR PRODOM; PD003649; Interleukin-2; 1.
DR SMART; SM00189; IL2; 1.
DR PROSITE; PS00424; INTERLEUKIN_2; 1.
KW Cytokine; Glycoprotein; Immune response; Signal; Growth factor;
KW T-cell.
FT CHAIN 1 20 INTERLEUKIN-2.
FT CHAIN 21 155 BY SIMILARITY.
FT CARBOHYD 79 127 O-LINKED (GALNAC... ) (BY SIMILARITY).
FT DISULFID 23 23
FT CONFLICT 3 5 RMQ -> QIP (IN REF. 2).
FT CONFLICT 22 22 P -> T (IN REF. 2).
FT CONFLICT 30 30 T -> P (IN REF. 2).
FT CONFLICT 51 51 L -> P (IN REF. 2).
FT CONFLICT 71 71 D -> A (IN REF. 2).
FT CONFLICT 89 89 D -> E (IN REF. 2).
FT CONFLICT 99 99 R -> L (IN REF. 2).
FT CONFLICT 107 113 YMASLKG -> SMDNIKR (IN REF. 2).
FT CONFLICT 140 140 Q -> L (IN REF. 2).
FT CONFLICT 144 144 T -> I (IN REF. 2).
FT CONFLICT 154 154 L -> M (IN REF. 2).
SQ SEQUENCE 155 AA; 17627 MW; 816667DFA052EDF CRC64;

Query Match 75.0%; Score 30; DB 1; Length 155;
Best Local Similarity 77.8%; Pred. No. 38;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 36 SLLLDLQL 44

RESULT 10
IL2_CAPHI STANDARD; PRT; 155 AA.
AC P36835; P79156;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Interleukin-2 precursor (IL-2) (T-cell growth factor) (TCGF).
GN IL2.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Caprinae; Capra.
```



```
OC Schizosacccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A., AND CHARACTERIZATION.
RC STRAIN=FY254;
RX MEDLINE=20062836; PubMed=10593913;
RA Labbe S., Pena M.O., Fernandes A.R., Thiele D.J.;
RT "A copper-sensing transcription factor regulates iron uptake genes in
Schizosaccharomyces pombe.";
RL J. Biol. Chem. 274:36252-36260(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Wood V., Rajandream M.A., Barrell B.G., Volkkaert G.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: REQUIRED FOR HIGH AFFINITY COPPER (PROBABLY REDUCED CU
CC 1) TRANSPORT INTO THE CELL.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- INDUCTION: BY COPPER DEPRIVATION, AND RERESSED BY COPPER
CC EFFICIENCY.
CC -1- SIMILARITY: BELONGS TO THE SLC31A FAMILY OF TRANSPORTERS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF175405; AAD51064.1; -.
DR EMBL; AJ243833; CAB52305.1; -.
DR EMBL; AL035592; CAB38165.1; -.
KW Transmembrane; Transport; Copper transport.
FT DOMAIN 1 142
FT TRANSFEM 143 163
FT DOMAIN 164 232
FT TRANSFEM 233 253
FT DOMAIN 254 289
FT DOMAIN 107 112
FT POLY-THR
SQ SEQUENCE 289 AA; 32020 MW; D6A07DF015B81322 CRC64;
Query Match 75.0%; Score 30; DB 1; Length 289;
Best Local Similarity 77.8%; Pred. No. 71;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLPLQL 9
Db 40 NLLPLQL 48
RESULT 14
CDB_HUMAN
ID CDB_HUMAN STANDARD; PRT; 333 AA.
AC P29016;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE T-cell surface glycoprotein CD1b precursor (CD1b antigen).
GN CD1B.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88097453; PubMed=2447586;
RA Martin L.H., Calabi F., Lefebvre F.-A., Bilsland C.A., Millstein C.;
RT "Structure and expression of the human thymocyte antigens CD1a, CD1b,
RT and CD1c.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:9189-9193(1987).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89341413; PubMed=2701945;
RA Aruffo A., Seed B.;
RT "Expression of cDNA clones encoding the thymocyte antigens CD1a, b, c
demonstrates a hierarchy of exclusion in fibroblasts.";
RL J. Immunol. 143:1723-1730(1989).
CC -1- FUNCTION: NOT KNOWN.
CC -1- SUBUNIT: ASSOCIATES NON-COVALENTLY WITH BETA-2-MICROGLOBULIN.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: EXPRESSED ON CORTICAL THYMOCYTES, ON CERTAIN
CC T-CELL LEUKEMIAS, AND IN VARIOUS OTHER TISSUES.
CC -1- SIMILARITY: BELONGS TO THE IMMNOGLOBULIN SUPERFAMILY.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD1 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd1.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22173; AAA51940.1; -.
DR EMBL; M22168; AAA51940.1; JOINED.
DR EMBL; M22169; AAA51940.1; JOINED.
DR EMBL; M22170; AAA51940.1; JOINED.
DR EMBL; M22171; AAA51940.1; JOINED.
DR EMBL; M22172; AAA51940.1; JOINED.
DR EMBL; M28826; AAA51939.1; -.
DR PIR; B39957; B39957.
DR PIR; B45801; B45801.
DR HSP; P11609; ICD1.
DR MIM; 188360; -.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003597; Ig_cl.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IGcl; 1.
KW Glycoprotein; Signal; Transmembrane; Immunoglobulin domain;
KW Multigene family.
FT SIGNAL 1 17
FT CHAIN 18 333
FT DOMAIN 18 303
FT TRANSFEM 304 324
FT DOMAIN 325 333
FT DISULFID 120 184
FT DISULFID 224 279
FT CARBOHYD 38 38
FT CARBOHYD 75 75
FT CARBOHYD 146 146
FT CARBOHYD 258 258
SQ SEQUENCE 333 AA; 36939 MW; 7D02B670D4DB8CC7 CRC64;
Query Match 75.0%; Score 30; DB 1; Length 333;
Best Local Similarity 75.0%; Pred. No. 81;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 LLLPLQL 9
Db 2 LLLPLQL 9
RESULT 15
PSPD_BOVIN
ID PSPD_BOVIN STANDARD; PRT; 369 AA.
AC P35246;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Pulmonary surfactant-associated protein D precursor (SP-D) (PSP-D).
GN SPFPD OR SPFP4.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 208-247.
 RC TISSUE=Lung;
 RX MEDLINE=93170856; PubMed=8436402;
 RA Lim B.L., Lu J., Reid K.B.M.;
 RT "Structural similarity between bovine conglutinin and bovine lung
 RT surfactant protein D and demonstration of liver as a site of
 RT synthesis of conglutinin.";
 RL Immunology 78:159-165(1993).
 CC -1- FUNCTION: CONTRIBUTES TO THE LUNG'S DEFENSE AGAINST INHALED
 CC MICROORGANISMS. BINDS STRONGLY MALTOSE RESIDUES AND TO A LESSER
 CC EXTENT OTHER ALPHA-GLUCOSYL MOETIES. IT COULD PARTICIPATE IN THE
 CC EXTRACELLULAR REORGANIZATION OR TURNOVER OF PULMONARY SURFACTANT.
 CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -1- MISCELLANEOUS: PULMONARY SURFACTANT
 CC PROTEIN. THERE ARE 4 SURFACTANT ASSOCIATED PROTEIN: 2 COLLAGENOUS,
 CC CARBOHYDRATE-BINDING GLYCOPROTEINS (SP-A AND SP-D) AND 2 SMALL
 CC HYDROPHOBIC PROTEINS (SP-B AND SP-C).
 CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

 DR EMBL: X75911; CAA53510.1; --
 DR PIR: S33603; S33603.
 DR HSP: P35247; I808.
 DR InterPro: IPR000087; Collagen.
 DR InterPro: IPR001304; Lectin_c.
 DR Pfam: PF01391; Collagen; 3.
 DR Pfam: PF00059; lectin_c; 1.
 DR SMART: SM00034; CLECT; 1.
 DR PROSITE: PS00615; C-TYPE_LECTIN_1; 1.
 DR PROSITE: PS50041; C-TYPE_LECTIN_2; 1.
 DR Glycoprotein: Calcium; Surface film; Gaseous exchange; Hydroxylation;
 KW Signal; Lectin; Collagen; Repeat; Coiled coil.
 FT SIGNAL 1 20
 FT CHAIN 21 369
 FT FT
 FT FT
 FT DOMAIN 46 216
 FT DOMAIN 217 248
 FT DOMAIN 273 369
 FT DISULFID 275 367
 FT DISULFID 345 359
 FT CARBOHYD 90 90
 FT MOD_RES 78 78
 FT MOD_RES 87 87
 FT MOD_RES 96 96
 FT MOD_RES 99 99
 FT MOD_RES 165 165
 FT MOD_RES 171 171
 FT SEQUENCE 369 AA; 37361 MW; 07D88B24E0AEB2E3 CRC64;
 Query Match 75.0%; Score 30; DB 1; Length 369;
 Best Local Similarity 75.0%; Pred. No. 90;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 QY 2 LLLPLQL 9
 Dd 2 LLLPLSVL 9
 RESULT 16
 CONG_BOVIN

ID CONG_BOVIN STANDARD; PRT; 371 AA.
 AC P23805;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Conglutinin precursor.
 GN CGN1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93213261; PubMed=8460993;
 RA Suzuki Y., Yin Y., Makino M., Kurimura T., Wakamiya N.;
 RT "Cloning and sequencing of a cDNA coding for bovine conglutinin.";
 RL Biochem. Biophys. Res. Commun. 191:335-342(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93277452; PubMed=7684896;
 RA Lu J., Laurson S.B., Thiel S., Jensenius J.C., Reid B.M.;
 RT "The cDNA cloning of conglutinin and identification of liver as a
 RT primary site of synthesis of conglutinin in members of the Bovidae.";
 RL Biochem. J. 292:157-162(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Liver;
 RX MEDLINE=94215917; PubMed=8163202;
 RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,
 RA Tauber A.I., Sastry K.N.;
 RT "Bovine conglutinin (BC) mRNA expressed in liver: cloning and
 RT characterization of the BC cDNA reveals strong homology to surfactant
 RT protein-D.";
 RL Gene 141:277-281(1994).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Semen;
 RX MEDLINE=94267222; PubMed=8207234;
 RA Liou L.S., Sastry R., Hartshorn K.L., Lee Y.M., Okarma T.B.,
 RA Tauber A.I., Sastry K.N.;
 RT "Bovine conglutinin gene exon structure reveals its evolutionary
 RT relationship to surfactant protein-D.";
 RL J. Immunol. 153:173-180(1994).
 RN [5]
 RP SEQUENCE OF 21-371.
 RX MEDLINE=91131556; PubMed=1993651;
 RA Lee Y.-M., Leiby K.R., Allar J., Paris K., Lerch B., Okarma T.B.;
 RT "Primary structure of bovine conglutinin, a member of the C-type
 RT animal lectin family.";
 RL J. Biol. Chem. 266:2715-2723(1991).
 RN [6]
 RP PRELIMINARY SEQUENCE OF 21-52.
 RX MEDLINE=87184551; PubMed=3566740;
 RA Young N.M., Leon M.A.;
 RT "The carbohydrate specificity of conglutinin and its homology to
 RT proteins in the hepatic lectin family.";
 RL Biochem. Biophys. Res. Commun. 143:645-651(1987).
 CC -1- FUNCTION: CALCIUM-DEPENDENT LECTIN-LIKE PROTEIN WHICH BINDS TO A
 CC YEAST CELL WALL EXTRACT & IMMUNE COMPLEXES THROUGH THE COMPLEMENT
 CC COMPONENT (C3BI). IT IS CAPABLE OF BINDING NONREDUCING TERMINAL
 CC N-ACETYLGLUCOSAMINE, MANNOSE, AND FUCOSE RESIDUES.
 CC -1- SUBUNIT: OLIGOMERIC COMPLEX OF 4 SET OF HOMOTRIMERS.
 CC -1- SIMILARITY: CONTAINS A COLLAGENOUS DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

```
CC EMBL; D14085; BAA03170.1; -
DR EMBL; X71774; CAA50665.1; -
DR EMBL; L18871; AAA20126.1; -
DR EMBL; U06860; AAB60624.1; -
DR EMBL; U06854; AAB60624.1; JOINED.
DR EMBL; U06855; AAB60624.1; JOINED.
DR EMBL; U06856; AAB60624.1; JOINED.
DR EMBL; U06857; AAB60624.1; JOINED.
DR EMBL; U06858; AAB60624.1; JOINED.
DR EMBL; U06859; AAB60624.1; JOINED.
DR PIR; A29416; A29416.
DR PIR; A23740; A23740.
DR HSSP; P35247; JN0450.
DR InterPro; IPR001304; Lectin_c.
DR Pfam; PF01391; Collagen; 3.
DR SMART; SM00034; CLECT; 1.
DR PROSITE; PS00615; C-TYPE LECTIN.1; 1.
DR PROSITE; PS50041; C-TYPE LECTIN.2; 1.
KW Lectin; Hydroxylation; Glycoprotein; Mannose-binding; Membrane;
KW Collagen; Repeat; Calcium; Signal.
FT SIGNAL 1 20
FT CHAIN 21 371
FT DOMAIN 46 216
FT DOMAIN 273 371
FT MOD_RES 63 63
FT MOD_RES 87 87
FT MOD_RES 99 99
FT MOD_RES 135 135
FT MOD_RES 141 141
FT MOD_RES 159 159
FT MOD_RES 162 162
FT MOD_RES 198 198
FT SITE 201 203
FT DISULFID 275 369
FT CARBOHYD 337 337
FT CONFLICT 173 173
FT CONFLICT 210 210
FT CONFLICT 218 218
FT CONFLICT 272 272
SQ SEQUENCE 371 AA; 37994 MW; 867BB41992544B1F CRC64;

Query Match 75.0%; Score 30; DB 1; Length 371;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
DB 2 LLLPLSVL 9

RESULT 17
ID YHHJ_ECOLI STANDARD; PRT; 374 AA.
AC P31953; P76703;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein yhhj.
GN YHHJ OR B3485.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;

-----
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RL region from 76.0 to 81.5 minutes."
RN Nucleic Acids Res. 22:2576-2586(1994).
[2]
RP SEQUENCE OF 225-374 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=93259920; PubMed=8387990;
RA Zhao S., Sandt C.H., Feulner G., Vlazny D.A., Gray J.A., Hill C.W.;
RT "Rhs elements of Escherichia coli K-12: complex composites of shared
and unique components that have different evolutionary histories."
RL J. Bacteriol. 175:2799-2808(1993).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane
(Potential).
CC -!- SIMILARITY: BELONGS TO THE ABC-2 SUBFAMILY OF INTEGRAL MEMBRANE
PROTEINS.
CC -!- SIMILARITY: TO E. COLI YBHR AND YBHS AND TO THE C-TERMINAL OF
E. COLI YHIG.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
CC -----
CC EMBL; U00039; AAB18460.1; ALT_INIT.
CC EMBL; AE000424; AAC76510.1; ALT_INIT.
CC EMBL; L02370; AAC61886.1; -.
CC EcoGene; EGI1767; yhhj.
CC InterPro; IPR000412; ABC2_transport.
CC PROSITE; PS00890; ABC2_MEMBRANE; 1.
KW Hypothetical protein; Transport; Transmembrane; Inner membrane;
Complete proteome.
FT TRANSMEM 23 43
FT TRANSMEM 173 193
FT TRANSMEM 230 250
FT TRANSMEM 256 276
FT TRANSMEM 284 304
FT TRANSMEM 343 363
FT TRANSMEM 374 AA; 41061 MW; 02895FB13F493391 CRC64;
SQ SEQUENCE 374 AA; 41061 MW; 02895FB13F493391 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 374;
Best Local Similarity 75.0%; Pred. No. 91;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
DB 294 VLLPLQL 301

RESULT 18
ID UL49_HSVSA STANDARD; PRT; 435 AA.
AC Q01046;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Gene 66 protein.
GN 66 OR ECUF6.
OS Herpesvirus saimiri (strain 11).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Rhadinovirus.
OX NCBI_TaxID=10383;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92333688; PubMed=1321287;
RA Albrecht J.-C., Nicholas J., Biller D., Cameron K.R., Biesinger B.,
RA Newman C., Wittmann S., Craxton M.A., Coleman H., Fleckenstein B.,
RA Honess R.W.;
RT "Primary structure of the herpesvirus saimiri genome."
RL J. Virol. 66:5047-5058(1992).
RN [2]
```

```

RP SEQUENCE FROM N.A.
RX MEDLINE=9220228; PubMed=1314457;
RA Nicholas J., Cameron K.R., Coleman H., Newman C., Honess R.W.;
RT "Analysis of nucleotide sequence of the rightmost 43 kbp of
RT herpesvirus saimiri (HVS) L-DNA: general conservation of genetic
RT organization between HVS and Epstein-Barr virus.";
RL Virology 188:296-310(1992).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS TOGETHER EBV BFRF2,
CC HVS-1 66, AND HCMV UL49.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X64346; CAA45689.1; -.
DR EMBL; M86409; AAA46142.1; -.
DR PIR; A36813; A36813.
DR InterPro; IPR004339; UL49.
DR Pfam; PF03117; UL49; 1.
DR Hypothetical protein.
SQ SEQUENCE 435 AA; 50126 MW; 79F0F834AD1B69D1 CRC64;
Query Match 75.0%; Score 30; DB 1; Length 435;
Best Local Similarity 87.5%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLPLQLI 8
DB 31 SLLPLQLI 38
III IIII

RESULT 19
YICE_ECOLI
ID YICE_ECOLI STANDARD; PRT; 463 AA.
AC P27432;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative purine permease yice.
GN YICE OR B3654 OR 25082 OR ECS4530.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=93315143; PubMed=7686882;
RA Burland V.D., Plunkett G. III, Daniels D.L., Blattner F.R.;
RT "DNA sequence and analysis of 136 kilobases of the Escherichia coli
RT genome: organizational symmetry around the origin of replication.";
RL Genomics 16:551-561(1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;

```

```

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
[4]
RP SEQUENCE OF 1-7 FROM N.A.
RC STRAIN=K12;
RX MEDLINE=91203811; PubMed=2017136;
RA Kalman M., Gentry D., Cashel M.;
RT "Characterization of the Escherichia coli K12 glts glutamate permease
RT gene.";
RL Mol. Gen. Genet. 225:379-386(1991).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE XANTHINE/URACIL PERMEASES FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L10328; AAA62007.1; -.
DR EMBL; AE000443; AAC76678.1; -.
DR EMBL; AE005593; AAG58799.1; -.
DR EMBL; AP002566; BAB37953.1; -.
DR EMBL; X17499; -. NOT_ANNOTATED_CDS.
DR EcoGene; EG11194; yice.
DR InterPro; IPR000444; Xan_ur_permease.
DR Pfam; PF00860; xan_ur_permease; 1.
DR PROSITE; PS01116; XANTH_URACIL_PERMASE; 1.
KW Hypothetical protein; Transmembrane; Transport; Complete proteome.
FT TRANSMEM 43 63 POTENTIAL.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 126 146 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 192 212 POTENTIAL.
FT TRANSMEM 222 242 POTENTIAL.
FT TRANSMEM 260 280 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 409 429 POTENTIAL.
FT TRANSMEM 439 459 POTENTIAL.
SQ SEQUENCE 463 AA; 48868 MW; 868718559103456E CRC64;
Query Match 75.0%; Score 30; DB 1; Length 463;
Best Local Similarity 66.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLPLQLI 9
DB 269 SLLPLMLV 277
IIIIII ::

RESULT 20
K2C4_MOUSE
ID K2C4_MOUSE STANDARD; PRT; 524 AA.
AC P07744;
DT 01-AUG-1988 (Rel. 08, Created)
DT 01-AUG-1988 (Rel. 08, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Keratin, type II cytoskeletal 4 (Cytokeratin 4) (Cytoskeletal 57 kDa
DE keratin).
DE KRT4 OR KRT2-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE-86120369; PubMed-2418416;
RA Knapp B., Rontrop M., Schweizer J., Winter H.;
RT "Nonepidermal members of the keratin multigene family: cDNA sequences
and in situ localization of the mRNAs.";
RL Nucleic Acids Res. 14:751-763(1986).
CC -1- SUBUNIT: HETEROTETRAMER OF TWO TYPE I AND TWO TYPE II KERATINS.
CC MICROFILAMENTAR KERATIN, I (ACIDIC) AND II (NEUTRAL TO BASIC)
(40-55 AND 56-70 KILODALTONS, RESPECTIVELY).
CC -1- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: X03491; CAA27207.1; -
CC PIR: A23518; A23518.
CC HSP: P02876; 2WGC.
CC MGD: MGI:96701; Krt2-4.
CC InterPro: IPR001664; IF.
CC Pfam: PF00038; filament_1.
CC PRINTS: PR01276; TYPE2KERATIN.
CC PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Keratin; Phosphorylation.
FT DOMAIN 1 145 HEAD.
FT DOMAIN 146 455 ROD.
FT DOMAIN 456 524 TAIL.
FT DOMAIN 146 181 COIL 1A.
FT DOMAIN 182 200 LINKER 1.
FT DOMAIN 201 292 COIL 1B.
FT DOMAIN 293 316 LINKER 12.
FT DOMAIN 317 455 COIL 2.
FT MOD_RES 51 51 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 524 AA; 56444 MW; 74F8E4B38BFB6B9 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 524;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQI 8
DB 126 SLLTPLQV 133

RESULT 21
PPBT_MOUSE
ID PPBT_MOUSE STANDARD; PRT; 524 AA.
AC P09242;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Alkaline phosphatase, tissue-nonspecific isozyme precursor
(EC 3.1.3.1) (AP-TNAP) (TNSALP).
GN ALPL OR AKP2 OR AKP-2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE-88041048; PubMed-3478679;
RA Terao M., Mintz B.;
RT "Cloning and characterization of a cDNA coding for mouse placental
alkaline phosphatase.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:7051-7055(1987).

[2]
RN SEQUENCE OF 1-6 FROM N.A.
RX MEDLINE-90330663; PubMed-2165496;
RA Brown N.A., Stofko R.E., Uhler M.D.;
RT "Induction of alkaline phosphatase in mouse L cells by overexpression
of the catalytic subunit of cAMP-dependent protein kinase.";
RL J. Biol. Chem. 265:13181-13189(1990).
CC -1- CATALYTIC ACTIVITY: An orthophosphoric monoester + H(2)O -> an
alcohol + phosphate.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- MISCELLANEOUS: IN MOST MAMMALS THERE ARE FOUR DIFFERENT ISOZYMES:
PLACENTAL, PLACENTAL-LIKE, INTESTINAL AND TISSUE NON-SPECIFIC
(LIVER/BONE/ KIDNEY).
CC -1- SIMILARITY: BELONGS TO THE ALKALINE PHOSPHATASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: J02980; AAA39928.1; -
CC EMBL: M54798; AAA37217.1; -
CC PIR: A28409; A28409.
CC HSP: P00634; 2ANH.
CC MGD: MGI:87983; Akp2.
CC InterPro: IPR001952; Alk_phosphatse.
CC Pfam: PF00245; alk_phosphatase; 1.
CC PRINTS: PR00113; ALKPHPTASE.
CC PRODOM: PD001868; Alk_phosphatse; 1.
CC SMART: SM00098; alkPPC; 1.
CC PROSITE: PS00123; ALKALINE_PHOSPHATASE; 1.
KW Hydrolase; Zinc; Magnesium; Phosphorylation; Transmembrane; Placenta;
KW Multigene family; Glycoprotein; GPI-anchor; Signal.
FT SIGNAL 1 17
FT CHAIN 18 524 ALKALINE PHOSPHATASE, TISSUE-NONSPECIFIC
FT MOD_RES 110 110 ISOZYME.
FT ACT_SITE 110 110 PHOSPHORYLATION.
FT CARBOHYD 140 140 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 303 303 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 430 430 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 524 AA; 57455 MW; 59D99113A60FA050 CRC64;

Query Match 75.0%; Score 30; DB 1; Length 524;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQIL 9
DB 510 ALLPLAVL 518

RESULT 22
COX1_MYCTU
ID COX1_MYCTU STANDARD; PRT; 573 AA.
AC O53290;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable cytochrome c oxidase polypeptide I (EC 1.9.3.1) (Cytochrome
AA3 subunit 1).
GN CTAD OR RV3043C OR MT3128 OR MTVD12.58C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=H37RV;
RX MEDLINE=98295987; PubMed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Teklaia F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: CYTOCHROME C OXIDASE IS THE COMPONENT OF THE RESPIRATORY
CC CHAIN THAT CATALYZES THE REDUCTION OF OXYGEN TO WATER. SUBUNITS 1-
CC 3 FORM THE FUNCTIONAL CORE OF THE ENZYME COMPLEX. CO I IS THE
CC CATALYTIC SUBUNIT OF THE ENZYME. ELECTRONS ORIGINATING IN
CC CYTOCHROME C ARE TRANSFERRED VIA THE COPPER A CENTER OF SUBUNIT 2
CC AND HEME A OF SUBUNIT 1 TO THE BIMETALLIC CENTER FORMED BY HEME A3
CC AND COPPER B (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: 4 ferrocyclochrome c + O(2) = 4 ferricytochrome
CC c + 2 H(2)O.
CC -!- PATHWAY: TERMINAL STEP IN THE RESPIRATORY CHAIN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE HEME-COPPER RESPIRATORY OXIDASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL021287; CAA16128.1; -
CC EMBL; AE007130; AAK47458.1; -
CC HSSP; P18401; 1FFT.
CC TIGR; MT3128; -
CC TuberculList; RV3043c; -
CC InterPro: IPR000883; COX1.
CC Pfam; PF00115; COX1; 1.
CC PRINTS; PR01165; CYCOXIDASE1.
CC PROSITE; PS00077; COX1; 1.
CC Oxidoreductase; Heme; Copper; Transmembrane; Respiratory chain;
KW Complete proteome.
FT TRANSMEM 40 60 POTENTIAL.
FT TRANSMEM 89 109 POTENTIAL.
FT TRANSMEM 121 141 POTENTIAL.
FT TRANSMEM 170 190 POTENTIAL.
FT TRANSMEM 213 233 POTENTIAL.
FT TRANSMEM 258 278 POTENTIAL.
FT TRANSMEM 290 310 POTENTIAL.
FT TRANSMEM 315 335 POTENTIAL.
FT TRANSMEM 359 379 POTENTIAL.
FT TRANSMEM 398 418 POTENTIAL.
FT TRANSMEM 433 453 POTENTIAL.
FT TRANSMEM 476 496 POTENTIAL.
FT METAL 86 86 IRON (HEME A) (PROBABLE).
FT METAL 264 264 COPPER B (PROBABLE).
FT METAL 268 268 COPPER B (PROBABLE).
FT METAL 313 313 COPPER B (PROBABLE).

FT METAL 314 314 COPPER B (PROBABLE).
FT METAL 397 397 IRON (HEME A3) (PROBABLE).
FT METAL 399 399 IRON (HEME A) (PROBABLE).
SQ SEQUENCE 573 AA; 63672 MW; 9B0F91666F482BEE CRC64;
Query Match 75.0%; Score 30; DB 1; Length 573;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 SLLPLQI-8
Db 104 NLVLPLQI 111
RESULT 23
NDCL_HUMAN STANDARD; PRT; 592 AA.
ID NDCL_HUMAN AC Q13183;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Renal sodium/dicarboxylate cotransporter (Na(+)/dicarboxylate
DE cotransporter).
DE GN SLC13A2 OR NADCL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=96199379; PubMed=8967342;
RA Pajor A.M.;
RT "Molecular cloning and functional expression of a sodium-dicarboxylate
RT cotransporter from human kidney.";
RL Am. J. Physiol. 270:F642-F648(1996).
CC -!- FUNCTION: COTRANSPORT OF SODIUM AND DICARBOXYLATES SUCH AS
CC SUCCINATE AND CITRATE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE NADC/P/PHO87 FAMILY OF TRANSPORTERS.
CC NADC SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U76209; AAA98504.1; -
CC MIM; 604148; -
CC InterPro: IPR001898; Na_sulph_symp.
CC Pfam; PF00939; Na_sulph_symp; 1.
CC PROSITE; PS01271; NA_SULFATE; 1.
CC Transprot; Transmembrane; Sodium transport; Symport.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 53 73 POTENTIAL.
FT TRANSMEM 86 106 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 221 241 POTENTIAL.
FT TRANSMEM 274 294 POTENTIAL.
FT TRANSMEM 324 344 POTENTIAL.
FT TRANSMEM 371 391 POTENTIAL.
FT TRANSMEM 450 470 POTENTIAL.
FT TRANSMEM 482 502 POTENTIAL.
FT TRANSMEM 511 531 POTENTIAL.
FT TRANSMEM 545 565 POTENTIAL.
SQ SEQUENCE 592 AA; 64410 MW; 41137D6621A0872A CRC64;
Query Match 75.0%; Score 30; DB 1; Length 592;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;


```

KW Transports; Iron transport; Transmembrane; Glycoprotein; Macrophage.
FT NON_TER 1 1
FT DOMAIN <1 14 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 15 32 POTENTIAL.
FT DOMAIN 33 43 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 44 64 POTENTIAL.
FT DOMAIN 65 80 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 81 102 POTENTIAL.
FT DOMAIN 103 110 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 111 >126 POTENTIAL.
FT NON_TER 126 126
SQ SEQUENCE 126 AA; 13881 MW; 79C3372CB08D1C32 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 126;
Best Local Similarity 66.7%; Pred. No. 48;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
   :||||:|
Db 50 SLLPLPAVL 58

RESULT 26
CD8A_PONPY STANDARD; PRT; 198 AA.
AC P30433;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein CD8 alpha chain precursor (T-lymphocyte
DE differentiation antigen T8/LEU-2).
GN CD8A.
OS Pongo pygmaeus (Orangutan).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Pongo.
OX NCBI_TaxID=9600;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92307742; PubMed=1612644;
RA Lawlor D.A., Parham P.;
RT "Structure of CD8 alpha and beta chains of the orangutan: novel
RT patterns of mRNA splicing encoding hingeless polypeptides.";
RL Immunogenetics 36:121-125(1992).
CC -1- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT
CC WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN
CC THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO
CC CLASS I MHC MOLECULES ALPHA-3 DOMAINS.
CC -1- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN
CC LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORMS HOMODIMERS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X60223; CAA42784.1; -.
CC HSPSP; P01732; 1CD8
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR003596; Ig_V.
CC Pfam; PF00047; Ig; 1.
CC SMART; SM00406; IGV; 1.
CC Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell; MHC I;
KW Signal.
KW SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 198 T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA
FT CHAIN.

```

```

FT DOMAIN 22 145 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 146 166 POTENTIAL.
FT DOMAIN 167 198 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 22 135 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 43 115 BY SIMILARITY.
SQ SEQUENCE 198 AA; 22099 MW; F3EC093EADB05561 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 198;
Best Local Similarity 66.7%; Pred. No. 76;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
   :||||:|
Db 7 ALLPLALL 15

RESULT 27
NU4M_PHOSU STANDARD; PRT; 208 AA.
ID NU4M_PHOSU
AC O21515;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 4 (EC 1.6.5.3) (Fragment).
GN MTND4 OR MT-ND4 OR ND4.
OS Phodopus sungorus (Striped hairy-footed hamster) (Djungarian hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Phodopus.
OX NCBI_TaxID=10044;
RN [1]
RP SEQUENCE FROM N.A.
RA Engel S.R., Hogan K.M., Taylor J.F., Davis S.K.;
RT "Molecular systematics and paleobiogeography of the South American
RT sigmodontine rodents.";
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U83804; AAB87228.1; -.
CC InterPro; IPR001750; Oxidored_q1.
CC InterPro; IPR000260; Oxidored_q5_N.
CC Pfam; PF00361; Oxidored_q1; 1..N.
CC Pfam; PF01059; Oxidored_q5_N; 1.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion.
FT NON_TER 208 208
SQ SEQUENCE 208 AA; 23773 MW; F95ED0116E7675C3 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 208;
Best Local Similarity 66.7%; Pred. No. 80;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
   :||||:|
Db 97 SLLTLQVL 105

RESULT 28
CD8A_HUMAN STANDARD; PRT; 235 AA.
ID CD8A_HUMAN
AC P01732;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein CD8 alpha chain precursor (T-lymphocyte

```

DE differentiation antigen T8/Leu-2).
GN CD8A OR MAL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85099337; PubMed=3871356;
RA Littman D.R., Thomas Y., Maddon P.J., Chess L., Axel R.;
RT "The isolation and sequence of the gene encoding T8: a molecule
RT defining functional classes of T lymphocytes.";
RL Cell 40:237-246(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86103103; PubMed=3936473;
RA Parnes J.R., Sizer K.C., Sukhatme V.P., Hunkapiller T.;
RT "Structure of Leu-2/T8 as deduced from the sequence of a cDNA clone.";
RL Behring Inst. Mitt. 77:48-55(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=85124610; PubMed=3918796;
RA Sukhatme V.P., Sizer K.C., Vollmer A.C., Hunkapiller T.,
RA Parnes J.R.;
RT "The T cell differentiation antigen Leu-2/T8 is homologous to
RT immunoglobulin and T cell receptor variable regions.";
RL Cell 40:591-597(1985).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=90035142; PubMed=2509342;
RA Nakayama K.-I., Tokito S., Okumura K., Nakauchi H.;
RT "Structure and expression of the gene encoding CD8 alpha chain (Leu-
RT 2/T8).";
RL Immunogenetics 30:393-397(1989).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=89215302; PubMed=2496167;
RA Norment A.M., Lonberg N., Lacy E., Littman D.R.;
RT "Alternatively spliced mRNA encodes a secreted form of human CD8
RT alpha. Characterization of the human CD8 alpha gene.";
RL J. Immunol. 142:3312-3319(1989).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (2.6 ANGSTROMS) OF 22-135.
RX MEDLINE=92191292; PubMed=1547508;
RA Leahy D.J., Axel R., Hendrickson W.A.;
RT "Crystal structure of a soluble form of the human T cell coreceptor
RT CD8 at 2.6-A resolution.";
RL Cell 68:1145-1162(1992).
CC -1- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT
CC WITH MHC CLASS I BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN
CC THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO
CC CLASS I MHC MOLECULES ALPHA-3 DOMAINS.
CC -1- SUBUNIT: IN GENERAL HETERO DIMER OF AN ALPHA AND A BETA CHAIN
CC LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORMS HOMODIMERS.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- ALTERNATIVE PRODUCTS: VARIOUS PATTERNS OF DIFFERENTIAL SPLICING
CC OF CD8 ALPHA TRANSCRIPTS INVOLVE EXCISION OF THE TRANSMEMBRANE OR
CC CYTOPLASMIC DOMAINS.
CC -1- PTM: ALL OF THE FIVE MOST CARBOXYL-TERMINAL CYSTEINES ARE USED TO
CC FORM INTER-CHAIN DISULFIDE BONDS IN DIMERS AND HIGHER MULTIMERS,
CC WHILE THE FOUR AMINO-TERMINAL CYSTEINES ARE NOT (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD8a entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd8alpha.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR EMBL: M26315; AAA79217.1;
DR EMBL: M26313; AAA79217.1; JOINED.
DR EMBL: M26314; AAA79217.1; JOINED.
DR EMBL: M12824; AAA61133.1;
DR EMBL: M12828; AAB04637.1;
DR EMBL: M27161; AAA59674.1;
DR PIR: A01999; RWHUT8.
DR PIR: A22824; A22824.
DR PIR: JP0105; JP0105.
DR PIR: A30604; A30604.
DR PIR: A45888; A45888.
DR PDB: 1CD8; 31-JAN-94.
DR MIM: 186910;
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; Phosphorylation;
KW T-cell; MHC I; Signal; 3D-structure; Alternative splicing.
FT SIGNAL 1 21
FT CHAIN 22 235 T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA
FT DOMAIN 22 182 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 183 203 POTENTIAL.
FT DOMAIN 204 235 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 22 135 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 43 115
FT STRAND 24 27
FT TURN 35 36
FT STRAND 39 45
FT STRAND 54 59
FT STRAND 68 73
FT STRAND 79 80
FT TURN 82 83
FT TURN 86 88
FT STRAND 89 94
FT TURN 95 96
FT STRAND 97 102
FT HELIX 107 109
FT STRAND 111 119
FT TURN 120 121
FT STRAND 122 125
FT STRAND 129 131
SQ SEQUENCE 235 AA; 25729 MW; FCCA29BAA73726BB CRC64;
Query Match 72.5%; Score 29; DB 1; Length 235;
Best Local Similarity 66.7%; Pred. No. 91;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLLPLOIL 9
DB :|||||:
DB 7 ALLPLALL 15
RESULT 29
CD8A_CANFA STANDARD; PRT; 239 AA.
ID CD8A_CANFA AC F33706;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE T-cell surface glycoprotein CD8 alpha chain precursor.
GN CD8A.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BEAGLE; TISSUE=Thymus;
RX MEDLINE=94378217; PubMed=8091416;
RA Gorman S.D., Frewin M.R., Cobbold S.P., Waldmann H.;

RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
alpha antigens.";
RL Tissue Antigens 43:184-188(1994).
CC -!- FUNCTION: IDENTIFIES CYTOTOXIC/SUPPRESSOR T-CELLS THAT INTERACT
CC WITH MHC CLASS II BEARING TARGETS. CD8 IS THOUGHT TO PLAY A ROLE IN
CC THE PROCESS OF T-CELL MEDIATED KILLING. CD8 ALPHA CHAINS BINDS TO
CC CLASS II MHC MOLECULES ALPHA-3 DOMAINS.
CC -!- SUBUNIT: IN GENERAL HETERODIMER OF AN ALPHA AND A BETA CHAIN
CC LINKED BY TWO DISULFIDE BONDS. CAN ALSO FORM HOMODIMERS.
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L14287; AAB02294.1; -.
CC HSSP; P01732; ICD8.
CC InterPro; IPR003006; Ig_MHC.
CC InterPro; IPR003596; Ig_V.
CC Pfam; PF00047; Ig; 1.
CC SMART; SM00406; IGV; 1.
CC Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell; MHC I;
KW Signal.
KW CHAIN 1 21 POTENTIAL.
FT SIGNAL 22 239 T-CELL SURFACE GLYCOPROTEIN CD8 ALPHA
FT CHAIN 22 239 CHAIN.
FT FT 22 186 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 22 186 POTENTIAL.
FT TRANSMEM 187 210 POTENTIAL.
FT DOMAIN 211 239 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 25 139 IG-LIKE V-TYPE DOMAIN.
FT DISULFID 46 119 BY SIMILARITY.
FT CARBOHYD 156 156 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 239 AA; 26036 MW; 1018579779A5CB7B CRC64;

Query Match 72.5%; Score 29; DB 1; Length 239;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 7 :|||||:|
ALLPLALL 15

RESULT 30
FLGH_AGR5 STANDARD; PRT; 239 AA.
ID FLGH_AGR5 STANDARD; PRT; 239 AA.
AC Q44342;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Flagellar L-ring protein precursor (Basal body L-ring protein).
GN FLGH.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305138; PubMed=9161424;
RA Deakin W.J., Furniss C.S., Parker V.E., Shaw C.H.;
RT "Isolation and characterisation of a linked cluster of genes from
RT Agrobacterium tumefaciens encoding proteins involved in flagellar
RT basal-body structure";
RL Gene 189:135-137(1997).
CC -!- FUNCTION: ASSEMBLES AROUND THE ROD TO FORM THE L-RING AND PROBABLY
CC PROTECTS THE MOTOR/BASAL BODY FROM SHEARING FORCES DURING
CC ROTATION.
CC -----

CC -!- SUBUNIT: THE BASAL BODY CONSTITUTES A MAJOR PORTION OF THE
CC CELLULAR ORGANELLE AND CONSISTS OF FOUR RINGS (L,P,S, AND M)
CC MOUNTED ON A CENTRAL ROD (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Attached to the outer membrane by a lipid
CC anchor (Probable).
CC -!- SIMILARITY: BELONGS TO THE FLGH FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39941; AAB68972.1; -.
CC EMBL; U95165; AAB1794.1; -.
CC InterPro; IPR00527; Flag_Lring.
CC Pfam; PF02107; FLGH; 1.
CC PRINTS; PR01008; FLGLRINGFLGH.
CC PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Flagella; Outer membrane; Lipoprotein; Signal.
FT SIGNAL 1 18 POTENTIAL.
FT CHAIN 19 239 FLAGELLAR L-RING PROTEIN.
FT LIPID 19 19 N-ACYL DIGLYCERIDE (PROBABLE).
SQ SEQUENCE 239 AA; 26010 MW; 92AD5C3F1EC2182E CRC64;

Query Match 72.5%; Score 29; DB 1; Length 239;
Best Local Similarity 66.7%; Pred. No. 92;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 8 ALLPLALL 16

RESULT 31
ATPL_ODOSI STANDARD; PRT; 242 AA.
ID ATPL_ODOSI STANDARD; PRT; 242 AA.
AC Q00825;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP synthase A chain precursor (EC 3.6.3.14) (Subunit IV).
GN ATP.
OS Odontella sinensis.
OG Chlorophyta.
OC Eukaryota; stramenopiles; Bacillariophyta; Coscinodiscophyceae;
OC Bidulphiophycidae; Eupodiscaceae; Eupodiscaceae; Odontella.
OX NCBI_TaxID=2839;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92219274; PubMed=1532839;
RA Pantic P.G., Strotmann H., Kowalik K.V.;
RT "Chloroplast atpase genes in the diatom Odontella sinensis reflect
RT cyanobacterial characters in structure and arrangement.";
RL J. Mol. Biol. 224:529-536(1992).
RN [2]
RP SEQUENCE FROM N.A.
RX Kowalik K.V., Stoebe B., Schaffran I., Kroth-Pancic P., Freier U.;
RT "The chloroplast genome of a chlorophyll a+c-containing alga,
RT Odontella sinensis";
RL Plant Mol. Biol. Rep. 13:336-342(1995).
CC -!- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
CC DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE.
CC -!- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CC CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
CC SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
CC HAS THREE MAIN SUBUNITS: A, B AND C.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Chloroplast
CC thylakoid membrane.
CC -!- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; X60752; CAA43152.1; -;
 DR EMBL; 267753; CAA91689.1; -;
 DR PIR; S23356; S23356.
 DR Mendel; 4855; ODOI; atp1.2.
 DR InterPro; IPR000568; ATP_synt_A.
 DR Pfam; PF001119; ATP_synt_A; 1.
 DR PRINTS; PR00123; ATPASEA.
 DR PROSITE; PS00449; ATPASEA; 1.
 KW Hydrogen ion transport; CF(0); Chloroplast; Transmembrane; Signal.
 FT SIGNAL 1 ?
 FT CHAIN 1 ? BY SIMILARITY.
 FT CHAIN 242 ATP SYNTHASE A CHAIN.
 FT TRANSMEM 36 55
 FT TRANSMEM 95 113 POTENTIAL.
 FT TRANSMEM 132 151 POTENTIAL.
 FT TRANSMEM 219 238 POTENTIAL.
 SQ SEQUENCE 242 AA; 26898 MW; 2B04B423FCCA2698 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 242;
 Best Local Similarity 75.0%; Pred. No. 93;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
 Db 168 LLLPINIL 175

RESULT 32

CD48_HUMAN STANDARD; PRT; 243 AA.
 AC P09326;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE B-lymphocyte activation marker BLAST-1 precursor (BCM1 surface
 DE antigen) (Leucocyte antigen MEM-102) (TCT.1) (Antigen CD48).
 GN CD48 OR BCM1 OR BLAST1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91153858; PubMed-1999351;
 RA Vaughan H.A., Henning M.M., Purcell D.F.J., McKenzie I.F.C.,
 RA Sandrin M.S.;
 RT "The isolation of cDNA clones for CD48.";
 RL Immunogenetics 33:113-117(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-91153857; PubMed-1999350;
 RA Korinek V., Stefanova I., Angelisova P., Hilbert I., Horejsi V.;
 RT "The human leucocyte antigen CD48 (MEM-102) is closely related to the
 RT activation marker Blast-1.";
 RL Immunogenetics 33:108-112(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-88111558; PubMed-2828034;
 RA Staunton D.E., Thorley-Lawson D.A.;
 RT "Molecular cloning of the lymphocyte activation marker Blast-1.";
 RL EMBO J. 6:3695-3701(1987).
 RN [4]
 RP SEQUENCE OF 1-27 FROM N.A.
 RX MEDLINE-91141511; PubMed-1847502;
 RA Fisher R.C., Thorley-Lawson D.A.;
 RT "Characterization of the Epstein-Barr virus-inducible gene encoding

RT the human leukocyte adhesion and activation antigen BLAST-1 (CD48).";
 RL Mol. Cell. Biol. 11:1614-1623(1991).
 RN [5]
 RP PARTIAL SEQUENCE.
 RX MEDLINE-91237281; PubMed-1827826;
 RA del Porto P., Mami-Chouaib F., Bruneau J.-M., Jitsukawa S., Dumas J.,
 RA Harnois M., Hercend T.;
 RT "TCT.1, a target molecule for gamma/delta T cells, is encoded by an
 RT immunoglobulin superfamily gene (Blast-1) located in the CD1 region
 RT of human chromosome 1.";
 RL J. Exp. Med. 173:1339-1344(1991).
 CC -!- FUNCTION: LIGAND FOR CD2. MIGHT FACILITATE INTERACTION BETWEEN
 CC ACTIVATED LYMPHOCYTES. PROBABLY INVOLVED IN REGULATING T-CELL
 CC ACTIVATION.
 CC -!- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
 CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -!- DATABASE: NAME-PROW; NOTE=CD guide CD48 entry;
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd48.htm".
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL; M59904; AA62834.1; -;
 DR EMBL; X06341; CAA29647.1; -;
 DR EMBL; M37766; AAA36211.1; -;
 DR EMBL; M63911; AAA35602.1; -;
 DR PIR; A27236; A27236.
 DR PIR; A53244; A53244.
 DR MIM; 109530; -;
 DR InterPro; IPR003599; Ig.
 DR InterPro; IPR003006; Ig_MHC.
 DR Pfam; PF00047; Ig; 1.
 DR SMART; SM00409; IG; 1.
 KW Immunoglobulin domain; B-cell; Glycoprotein; GPI-anchor; T-cell;
 KW Signal.

FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 220 B-LYMPHOCYTE ACTIVATION MARKER BLAST-1.
 FT PROPEP 221 243 REMOVED IN MATURE FORM.
 FT LIPID 220 220 GPI-ANCHOR.
 FT DISULFID 154 196 POTENTIAL.
 FT CARBOHYD 40 40 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 44 44 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 104 104 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 162 162 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 189 189 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 2 2 C -> W (IN REF. 1).
 FT CONFLICT 66 66 I -> N (IN REF. 3).
 SQ SEQUENCE 243 AA; 27683 MW; 9E46D76FC36A512C CRC64;

Query Match 72.5%; Score 29; DB 1; Length 243;
 Best Local Similarity 75.0%; Pred. No. 94;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
 Db 15 LLLPLSL 22

RESULT 33

TRUB_NEIMA STANDARD; PRT; 306 AA.
 ID TRUB_NEIMA
 AC Q9JTX5;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE tRNA pseudouridine synthase B (BC 4.2.1.70) (tRNA pseudouridine 55

```

DE synthase) (Psi55 synthase) (Pseudouridylylate synthase) (Uracil
DE TRUB OR NMA1588.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
CC -1- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL162756; CAB84815.1; -
CC InterPro: IPR002501; TruB_N.
CC Pfam: PF01509; TruB_N; 1.
CC Lyase; tRNA processing; Complete proteome.
CC SEQUENCE 306 AA; 33515 MW; D066EDF8962977FB CRC64;

Query Match 72.5%; Score 29; DB 1; Length 306;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 228 SLLPLCDVL 236
|||||:|

RESULT 34
TRUB_NEIMB STANDARD; PRT; 307 AA.
AC Q8JYV1:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE tRNA pseudouridine synthase B (EC 4.2.1.70) (tRNA pseudouridine 55
DE synthase) (Psi55 synthase) (Pseudouridylylate synthase) (Uracil
DE hydrolyase)
GN TRUB OR NMB1374.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gunn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Cifton H., Clark E.B.,
RA Cotton M.D., Unterback T.R., Khouri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Massignani V., Pizza M., Grandi G., Sun L.,

```

```

RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58."
RL Science 287:1809-1815(2000).
CC -1- FUNCTION: FORMATION OF PSEUDOURIDINE AT POSITION 55 IN THE
CC PSI GC LOOP OF TRANSFER RNAs (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Uracil + D-ribose 5-phosphate = pseudouridine
CC 5'-phosphate + H(2)O.
CC -1- SIMILARITY: BELONGS TO THE TRUB FAMILY OF PSEUDOURIDINE SYNTHASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE002485; AAF41748.1; -
CC TIGR: NMB1374; -
CC InterPro: IPR002501; TruB_N.
CC Pfam: PF01509; TruB_N; 1.
CC Lyase; tRNA processing; Complete proteome.
CC SEQUENCE 307 AA; 33632 MW; 4B0FBADACEB99E79 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 307;
Best Local Similarity 66.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 229 SLLPLCDVL 237
|||||:|

RESULT 35
O59A_DROME STANDARD; PRT; 397 AA.
AC P81923: Q9UGV1:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Odorant receptor 59a.
GN OR59A OR OR59D.1 OR DOR59D.1 OR DOR46 OR AN6 OR CG9820.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A., AND TISSUE SPECIFICITY.
RX MEDLINE=99389723; PubMed=10458908;
Gao Q., Chess A.;
RT "Identification of candidate Drosophila olfactory receptors from
RT genomic DNA sequence."
RL Genomics 60:31-39(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

```


RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Markulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirska R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassaran D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*";
RL Science 287:2185-2195(2000).
[3]
RN IDENTIFICATION, AND TISSUE SPECIFICITY.
RP MEDLINE-99166868; PubMed-10069338;
RX Clyne P.J., Warr C.G., Freeman M.R., Lessing D., Kim J., Carlson J.R.;
RA "A novel family of divergent seven-transmembrane proteins: candidate
RT odorant receptors in *Drosophila*";
RL Neuron 22:327-338(1999).
[4]
RN IDENTIFICATION.
RP STRAIN-OREGON-R: TISSUE=Maxillary palps;
RX MEDLINE-99189757; PubMed-10089887;
RA Vossball L.B., Amrein H., Morozov P.S., Rhetsky A., Axel R.;
RT "A spatial map of olfactory receptor expression in the *Drosophila*
NT antenna";
RL Cell 96:725-736(1999).
CC -1- FUNCTION: PROBABLE ROLE IN THE ODORANT RESPONSE, BEING AN ODORANT
CC RECEPTOR.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: EXPRESSED IN NEURONS OF THE THIRD ANTENNAL
CC SEGMENT.
CC -1- SIMILARITY: BELONGS TO FAMILY DR-OR OF G-PROTEIN COUPLED
CC RECEPTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE003460; AAF47004.1; ALT-SEQ.
DR FlyBase; FBgn026384; Or59a.
DR InterPro; IPR004117; 7tm_6.
DR Pfam; PF02949; 7tm_6; 1.
KW Transmembrane; G-protein coupled receptor; Olfaction; Glycoprotein;
KW Multigene family.
FT DOMAIN 1 36 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 37 57 1 (POTENTIAL).
FT DOMAIN 58 68 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 69 92 2 (POTENTIAL).
FT DOMAIN 93 128 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 129 149 3 (POTENTIAL).
FT DOMAIN 150 179 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 180 200 4 (POTENTIAL).
FT DOMAIN 201 274 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 275 295 5 (POTENTIAL).
FT DOMAIN 296 301 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 302 322 6 (POTENTIAL).

FT DOMAIN 323 372 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 373 393 7 (POTENTIAL).
FT DOMAIN 394 397 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 397 AA; 46451 MW; 322E25A73288AC92 CRC64;
Query Match 72.5%; Score 29; DB 1; Length 397;
Best Local Similarity 75.0%; Pred. No. 1.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLPLPQI 8
Db 308 SLAMPLQI 315
RESULT 36
SECY_BORBU STANDARD; PRT; 434 AA.
AC OS1451;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Preprotein translocase secy subunit.
GN SECY OR BB0498.
OS *Borrelia burgdorferi* (Lyme disease spirochete).
OC Bacteria; Spirochaetales; Spirochaetaceae; *Borrelia*.
OX NCBI_TaxID=139;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35210 / B31;
RX MEDLINE-98065943; PubMed-9403685;
RA Fraser C.M., Casjens S., Huang W.M., Sutton G.G., Clayton R.A.,
RA Lathigra R., White O., Ketchum K.A., Dodson R., Hickey E.K., Gwinn M.,
RA Dougherty B., Tomb J.-F., Fleischmann R.D., Richardson D.,
RA Peterson J., Kerlavage A.R., Quackenbush J., Salzberg S., Hanson M.,
RA Ueberbuck R., Palmer N., Adams M.D., Gocayne J.D., Weidman J.,
RA Venter V.T., Matthey L., McDonald L., Artlich P., Bowman C.,
RA Garland S., Fujii C., Cotton M.D., Horst K., Roberts K., Hatch B.,
RA Smith H.O., Venter J.C.;
RT "Genomic sequence of a Lyme disease spirochete, *Borrelia burgdorferi*."
RL Nature 390:580-586(1997).
CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. INTERACTS WITH SECA AND SECE
CC TO ALLOW THE TRANSLLOCATION OF PROTEINS ACROSS THE PLASMA MEMBRANE.
CC BY FORMING PART OF A CHANNEL.
CC -1- SUBUNIT: ONE OF SEVEN SECRETORY PROTEINS (SECA-F & SECY) THAT
CC COMPOSE THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE SECY/SEC61-ALPHA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE001152; AAC66844.1; -.
DR TIGR; BB0498; -.
DR InterPro; IPR002208; Secy.
DR Pfam; PF00344; secy; 1.
DR PRINTS; PR00303; SECYTRNLCASE.
DR PROSITE; PS00755; SECY_1; 1.
DR PROSITE; PS00756; SECY_2; 1.
KW Protein transport; Translocation; Transmembrane; Complete proteome.
FT TRANSMEM 19 39 POTENTIAL.
FT TRANSMEM 73 93 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 148 168 POTENTIAL.
FT TRANSMEM 179 199 POTENTIAL.
FT TRANSMEM 209 229 POTENTIAL.
FT TRANSMEM 264 284 POTENTIAL.


```
FT TRANSMEM 300 320 POTENTIAL.
FT TRANSMEM 362 382 POTENTIAL.
FT TRANSMEM 391 411 POTENTIAL.
SQ SEQUENCE 434 AA; 48610 MW; 89FC03AC2B69BD7C CRC64;

Query Match 72.5%; Score 29; DB 1; Length 434;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
Db 274 ITPPLQL 281

RESULT 37
YA06_HUMAN STANDARD; PRT; 435 AA.
AC O60813;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical protein DJ845024.5 (Fragment).
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Grafham D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Rhodes S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MAPE FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AL022101; CAA17880.1; -
DR EMBL; AL045680; CAB41252.1; -
KW Hypothetical protein.
FT NON_TER 435 435
SQ SEQUENCE 435 AA; 50539 MW; 3D19443032BBB494 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 435;
Best Local Similarity 85.7%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLLPLQL 9
Db 298 LVPLQL 304

RESULT 38
PGLR_PRAE STANDARD; PRT; 462 AA.
AC Q02056;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Polygalacturonase precursor (EC 3.2.1.15) (PG) (Pectinase).
DE Persa americana (Avocado).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Laurales; Lauraceae; Persea.
OX NCBI_TaxID=3435;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC STRAIN=CV, HASS; TISSUE=Pericarp;
RX MEDLINE=93184201; PubMed=8095163;
RA Dopico B., Lowe A.L., Wilson I.D., Merodio C., Grierson D.;
RT "Cloning and characterization of avocado fruit mRNAs and their
RT expression during ripening and low-temperature storage.";
RL Plant Mol. Biol. 21:437-449(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, HASS; TISSUE=Mesocarp;
RX MEDLINE=94269193; PubMed=8208850;
RA Kutsuna S.Y., Lin A.C., Percival F.W., Laties G.G.,
RA Christoffersen R.E.;
RT "Ripening-related polygalacturonase cDNA from avocado.";
RL Plant Physiol. 103:289-290(1993).
CC -!- FUNCTION: ACTS IN CONCEPT WITH THE PECTINESTERASE, IN THE RIPENING
CC PROCESS. IS INVOLVED IN CELL WALL METABOLISM, SPECIFICALLY IN
CC POLYURONIDE DEGRADATION.
CC -!- CATALYTIC ACTIVITY: Random hydrolysis of 1,4-alpha-D-
CC galactosiduronic linkages in pectate and other galacturonans.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- DEVELOPMENTAL STAGE: IN RIPENING FRUIT.
CC -!- SIMILARITY: BELONGS TO FAMILY 28 OF GLYCOSYL HYDROLASES
CC (POLYGLACTURONASES).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X66426; CAA47055.1; -
DR EMBL; L06094; AAA32914.1; -
DR PIR; S28072; S28072.
DR InterPro; IPR000743; Polygalacturonase.
DR Pfam; PF00295; Glyco_hydro_28; 1.
DR PROSITE; PS00502; POLYGLACTURONASE; 1.
KW Hydrolase; Glycosidase; Cell wall; Signal; Fruit ripening;
KW Glycoprotein.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 462 POLYGLACTURONASE.
FT ACT_SITE 301 301 BY SIMILARITY.
FT CARBOHYD 294 294 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CONFLICT 182 182 S -> I (IN REF. 2).
FT CONFLICT 416 419 YIVG -> ILLE (IN REF. 2).
FT CONFLICT 442 462 LLREGLSFLFMKRRVHECSY -> TTEGKVYPSPCL
FT (IN REF. 2).
SQ SEQUENCE 462 AA; 50290 MW; ABCF068ACCB17A88 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 462;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
Db 6 LLLPISIL 13

RESULT 39
PTGB_BUCAI STANDARD; PRT; 477 AA.
AC P57437;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE pns system, glucose-specific IIBC component (EIIBC-GLC) (Glucose-
DE permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE (EC 2.7.1.69) (EIIG-GLC).
GN PTSG OR BU356.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
OS symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
```

```

OX NCBI_TaxID=118099;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids
RL Buchnera sp. APS.";
RL Nature 407:81-86(2000).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
CC -TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
CC AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
CC PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
CC PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
CC THE SUGAR (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Protein N-phosphohistidine + sugar = protein
CC histidine + sugar phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AP001119; BABI3060.1; ALT_INIT.
CC InterPro: IPR001996; PTS_EIIB.
CC InterPro: IPR003352; PTS_EIIC.
CC Pfam: PF00367; PTS_EIIB; 1.
CC Pfam: PF02378; PTS_EIIC; 1.
CC PROSITE: PS01035; PTS_EIIB_CVS; 1.
CC Phosphotransferase system; Sugar transport; Transferase;
CC Phosphorylation; Transmembrane; Complete proteome.
CC KW DOMAIN 1 390 EIIC DOMAIN.
CC FT DOMAIN 391 477 EIIB DOMAIN.
CC FT MOD_RES 421 421 PHOSPHORYLATION (BY SIMILARITY).
CC FT TRANSMEM 15 35 POTENTIAL.
CC FT TRANSMEM 51 71 POTENTIAL.
CC FT TRANSMEM 76 96 POTENTIAL.
CC FT TRANSMEM 112 132 POTENTIAL.
CC FT TRANSMEM 152 172 POTENTIAL.
CC FT TRANSMEM 191 211 POTENTIAL.
CC FT TRANSMEM 250 270 POTENTIAL.
CC FT TRANSMEM 280 300 POTENTIAL.
CC FT TRANSMEM 304 324 POTENTIAL.
CC FT TRANSMEM 357 377 POTENTIAL.
CC FT TRANSMEM 440 460 POTENTIAL.
CC SQ SEQUENCE 477 AA; 51727 MW; 5204C68E94477C5F CRC64;

Query Match 72.5%; Score 29; DB 1; Length 477;
Best Local Similarity 55.6%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLPLPLQL 9
||:|:|:|
Db 15 SLMLPVSVL 23

RESULT 40
PTGB_ECOLI
ID PTGB_ECOLI STANDARD; PRT; 477 AA.
AC P05053;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE PTS system, glucose-specific IIBC component (EIIBC-CLC) (Glucose-
DE perasease IIBC component) (Phosphotransferase enzyme II, BC component)
DE (EC 2.7.1.69) (EIIB-GLC).

```

```

GN PTSG OR GLCA OR UMG OR BIL01 OR Z1740 OR ECS1479.
OS Escherichia coli, and
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562, 83334;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87057327; PubMed=3023349;
RA Erni B., Zanolari B.;
RT "Glucose-permease of the bacterial phosphotransferase system. Gene
RL cloning, overproduction, and amino acid sequence of enzyme IIGlc.";
RL J. Biol. Chem. 261:16398-16403(1986).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:12453-1474(1997).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Aiba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajinaka M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horiuchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:11-22(2001).
RN [6]
RP PHOSPHORYLATION OF CVS-421.
RX MEDLINE=93280116; PubMed=8505292;
RA Meins M., Jenoe P., Mueller D., Richter W.J., Rosenbusch J.P.,
RA Erni B.;
RT "Cysteine phosphorylation of the glucose transporter of Escherichia
RT coli.";
RL J. Biol. Chem. 268:11604-11609(1993).
RN [7]
RP STRUCTURE BY NMR Of 391-476.
RX MEDLINE=94155901; PubMed=8112346;
RA Colic Gradolink S., Eberstadt M., Gemmecker G., Kessler H., Buhr A.,
RA Erni B.;
RT "The glucose transporter of Escherichia coli. Assignment of the 1H,

```

RT 13C and 15N resonances and identification of the secondary structure
of the soluble IIB domain.";
RL Eur. J. Biochem. 219:945-952(1994).
RN [8].
RP STRUCTURE BY NMR OF 386-477.
RX MEDLINE-96378619; PubMed-8784182;
RA Eberstadt M., Golic Grdadolnik S., Gemmecker G., Kessler H., Buhr A.,
Erni B.;
RT "Solution structure of the IIB domain of the glucose transporter of
Escherichia coli".
RL Biochemistry 35:11286-11292(1996).
RN [9].
RP STRUCTURE BY NMR OF 386-477.
RX MEDLINE-97344161; PubMed-9200688;
RA Gemmecker G., Eberstadt M., Buhr A., Lanz R., Golic Grdadolnik S.,
Kessler H., Erni B.;
RT "Glucose transporter of Escherichia coli: NMR characterization of the
phosphotyrosine form of the IIB(Glc) domain and its binding interface
with the IIA(Glc) subunit".
RL Biochemistry 36:7408-7417(1997).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
THE SUGAR. THIS ENZYME IS ALSO A CHEMORECEPTOR MONITORING THE
ENVIRONMENT FOR CHANGES IN SUGAR CONCENTRATION.
CC -1- CATALYTIC ACTIVITY: protein N-phosphohistidine + sugar = protein
histidine + sugar phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: J02618; AAA24437.1; -
DR EMBL: AE000210; AAC74185.1; -
DR EMBL: D90745; BAA35908.1; -
DR EMBL: D90745; BAA35916.1; -
DR EMBL: AE005320; AAG55847.1; -
DR EMBL: AF002555; BAB34902.1; -
DR PIR; A25336; WQEC2G.
DR PDB; 1IBA; 23-MAR-96.
DR ECO2DBASE; H042.6; 6TH EDITION.
DR ECO2DBASE; H042.7; 6TH EDITION.
DR EcoGene; EGI0787; ptsG.
DR InterPro; IPR001996; PTS_EIIB.
DR InterPro; IPR003352; PTS_EIIC.
DR Pfam; PF00367; PTS_EIIB; 1.
DR Pfam; PF02378; PTS_EIIC; 1.
DR ProDom; PD001476; PTS_EIIB; 1.
DR ProSITE; PS01035; PTS_EIIB_CYS; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane; Inner membrane; 3D-structure;
Complete proteome.
KW DOMAIN 1 390 EIIC DOMAIN.
FT DOMAIN 391 477 EIIB DOMAIN.
FT MOD_RES 421 421 PHOSPHORYLATION.
FT TRANSMEM 15 35 POTENTIAL.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 152 172 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 250 270 POTENTIAL.
FT TRANSMEM 280 300 POTENTIAL.

FT TRANSMEM 310 330 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
FT TRANSMEM 443 463 POTENTIAL.
SQ SEQUENCE 477 AA; 50676 MW; D97A80FD64B74F73 CRC64;
Query Match 72.5%; Score 29; DB 1; Length 477;
Best Local Similarity 55.6%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLLPLOIL 9
DB 15 SLMLPVSVL 23
RESULT 41
PTGB_SALTY STANDARD; PRT; 477 AA.
AC P37439;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE PTS system, glucose-specific IIBC component (EIIBC-GLC) (Glucose-
permease IIBC component) (Phosphotransferase enzyme II, BC component)
DE (EC 2.7.1.69) (EIIC-GLC).
DE PTSG OR STM1203.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SB3507;
RA Stolz B., Wehrli C., Erni B.;
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE-21534948; PubMed-11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
LT2".
RL Nature 413:852-856(2001).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
-TRANSPORT SYSTEM. THE IICD DOMAINS CONTAIN THE SUGAR BINDING SITE
AND THE TRANSMEMBRANE CHANNEL; THE IIA DOMAIN CONTAINS THE PRIMARY
PHOSPHORYLATION SITE (THE DONOR IS PHOSPHO-HPR); IIA TRANSFERS ITS
PHOSPHORYL GROUP TO THE IIB DOMAIN WHICH FINALLY TRANSFERS IT TO
THE SUGAR.
CC -1- CATALYTIC ACTIVITY: protein N-phosphohistidine + sugar = protein
histidine + sugar phosphate.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Inner membrane.
CC -1- MISCELLANEOUS: THIS ENZYME IS ALSO A CHEMORECEPTOR MONITORING THE
ENVIRONMENT FOR CHANGES IN SUGAR CONCENTRATION.
CC -1- SIMILARITY: CONTAINS A PTS EIIB DOMAIN.
CC -1- SIMILARITY: CONTAINS A PTS EIIC DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
CC EMBL: X74629; CAA52702.1; -
DR EMBL: AE008752; AAL20132.1; -
DR PIR; S36620; S36620.
DR HSSP; P05053; 1IBA.

DR StyGene: SG10316; ptsg.
DR InterPro: IPR001996; PTS_EIIB.
DR InterPro: IPR003352; PTS_EIIC.
DR Pfam: PF00367; PTS_EIIB; 1.
DR Pfam: PF02378; PTS_EIIC; 1.
DR ProDom: PD001476; PTS_EIIB; 1.
DR ProSITE: PS01035; PTS_EIIB_CYS; 1.
KW Phosphotransferase system; Sugar transport; Transferase;
KW Phosphorylation; Transmembrane; Inner membrane; Complete proteome.
FT DOMAIN 1 390
FT DOMAIN 391 477
FT MOD_RES 421 421
FT TRANSMEM 15 35
FT TRANSMEM 51 71
FT TRANSMEM 80 100
FT TRANSMEM 112 132
FT TRANSMEM 152 172
FT TRANSMEM 191 211
FT TRANSMEM 250 270
FT TRANSMEM 280 300
FT TRANSMEM 310 330
FT TRANSMEM 356 376
FT TRANSMEM 443 463
FT CONFLICT 218 218 F -> S (IN REF. 1).
FT CONFLICT 379 379 L -> F (IN REF. 1).
FT CONFLICT 477 477 S -> H (IN REF. 1).
SQ SEQUENCE 477 AA; 50497 MW; 8CD10142542FF797 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 477;
Best Local Similarity 55.8%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 15 SLMLPVSVL 23

RESULT 42
YA04_HUMAN
ID AC O60810; STANDARD; PRT; 500 AA.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical protein DJ845024.2 (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Grafham D.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Rhodes S.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO THE MAPE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL022101; CAAL17877.1; -.
DR EMBL; AL049681; CAB41253.1; -.
KW Hypothetical protein.
FT NON_TER 1 1
FT NON_TER 500 500
SQ SEQUENCE 500 AA; 56247 MW; 1895CD8A8F14B7C3 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 500;
Best Local Similarity 85.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LPLQL 9
Db 363 LVLPLQL 369

RESULT 43
NRML_PIG

ID NRML_PIG STANDARD; PRT; 538 AA.
AC O77741; Q9NLY2;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Natural resistance-associated macrophage protein 1 (NRAMP 1).
GN SLC11A1 OR NRAMP1.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=97179291; PubMed=9027577;
RA Tuggie C.K., Schmitz C.B., Gingerich-Feil D.;
RT "Rapid communication: cloning of a pig full-length natural resistance
RL J. Anim. Sci. 75:277-277(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20143710; PubMed=10678911;
RA Zhang G., Wu H., Ross C.R., Minton E., Blecha F.;
RT "Cloning of porcine NRAMP1 and its induction by lipopolysaccharide,
RT tumor necrosis factor alpha, and interleukin-1beta: role of CD14 and
RT mitogen-activated protein kinases";
RL Infect. Immun. 68:1086-1093(2000).
CC -!- FUNCTION: DIVALENT TRANSITION METAL (IRON AND MANGANESE)
CC TRANSPORTER INVOLVED IN IRON METABOLISM AND HOST RESISTANCE TO
CC CERTAIN PATHOGENS. MACROPHAGE-SPECIFIC MEMBRANE TRANSPORT
CC FUNCTION. CONTROLS NATURAL RESISTANCE TO INFECTION WITH
CC INTRACELLULAR PARASITES PATHOGEN RESISTANCE INVOLVES SEQUESTRATION
CC OF FE2+ AND MN2+ COFACTORS OF BOTH PROKARYOTIC AND EUKARYOTIC
CC CATALASES AND SUPEROXIDE DISMUTASES, NOT ONLY TO PROTECT THE
CC MACROPHAGE AGAINST ITS OWN GENERATION OF REACTIVE OXYGEN SPECIES,
CC BUT TO DENY THE CATIONS TO THE PATHOGEN FOR SYNTHESIS OF ITS
CC PROTECTIVE ENZYMES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -!- SIMILARITY: BELONGS TO THE NRAMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U55068; AAC24491.1; -.
DR EMBL; AF132037; AAF36527.1; -.
DR InterPro: IPR001046; Nramp.
DR Pfam: PF01566; Nramp; 1.
DR PRINTS; PR00447; NATRESASSCMP.
DR ProDom; PD001861; Nramp; 1.
KW Transport; Iron transport; Transmembrane; Glycoprotein; Macrophage.
FT DOMAIN 1 54
FT TRANSMEM 55 75
FT TRANSMEM 76 81
FT TRANSMEM 82 102
FT TRANSMEM 103 139
FT TRANSMEM 140 160

```
FT DOMAIN 161 164 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 165 185 POTENTIAL.
FT DOMAIN 186 193 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 194 214 POTENTIAL.
FT DOMAIN 215 240 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 241 261 POTENTIAL.
FT DOMAIN 262 286 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 287 307 POTENTIAL.
FT DOMAIN 308 346 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 347 367 POTENTIAL.
FT DOMAIN 368 394 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 395 415 POTENTIAL.
FT DOMAIN 416 432 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 433 453 POTENTIAL.
FT DOMAIN 454 468 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 469 489 POTENTIAL.
FT DOMAIN 490 492 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 493 513 POTENTIAL.
FT DOMAIN 514 538 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 321 321 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 2 2 T -> S (IN REF. 2).
FT CONFLICT 6 6 D -> G (IN REF. 2).
FT CONFLICT 137 137 N -> T (IN REF. 2).
FT CONFLICT 164 164 R -> G (IN REF. 2).
FT CONFLICT 185 185 N -> D (IN REF. 2).
SQ SEQUENCE 538 AA; 55012 MW; 235DE941CDEB2E5B CRC64;

Query Match 72.5%; Score 29; DB 1; Length 538;
Best Local Similarity 66.7%; Pred. No. 2.le+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLPQIL 9
DB 433 SLLPFAVL 441

RESULT 44
YA94_HUMAN STANDARD; PRT; 538 AA.
AC Q9UP08;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DE Hypothetical protein KIAA1094.
GN KIAA1094.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99397452; PubMed=10470851;
RA Kikuno R., Nagase T., Ishikawa K.-I., Hirotsawa M., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XIV.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro."
RL DNA Res. 6:197-205(1999).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- TISSUE SPECIFICITY: UBQUITOUS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB029017; BAA83046.1; -.
CC Hypothetical protein; Transmembrane.
KW
```

```
FT TRANSMEM 14 34 POTENTIAL..
FT TRANSMEM 75 95 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT TRANSMEM 189 209 POTENTIAL.
FT TRANSMEM 225 245 POTENTIAL.
FT TRANSMEM 255 275 POTENTIAL.
FT TRANSMEM 298 318 POTENTIAL.
FT TRANSMEM 352 372 POTENTIAL.
FT TRANSMEM 402 422 POTENTIAL.
FT TRANSMEM 437 457 POTENTIAL.
FT TRANSMEM 473 493 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
SQ SEQUENCE 538 AA; 59267 MW; EB7D1BABD45362AD CRC64;

Query Match 72.5%; Score 29; DB 1; Length 538;
Best Local Similarity 66.7%; Pred. No. 2.le+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLPQIL 9
DB 525 SLLPLPYLL 533

RESULT 45
NR01_BISBI STANDARD; PRT; 548 AA.
ID NR01_BISBI
DT Q95102;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Natural resistance-associated macrophage protein 1 (NRAMP 1).
GN SLC11A1 OR NRAMP1.
OS Bison bison (American bison).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bison.
OX NCBI_TaxID=9901;
RN [1]
RP SEQUENCE FROM N.A.
RA Feng J., Li Y., Templeton J.W.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: DIVALENT TRANSITION METAL (IRON AND MANGANESE)
CC TRANSPORTER INVOLVED IN IRON METABOLISM AND HOST RESISTANCE TO
CC CERTAIN PATHOGENS. MACROPHAGE-SPECIFIC MEMBRANE TRANSPORT
CC FUNCTION. CONTROLS NATURAL RESISTANCE TO INFECTION WITH
CC INTRACELLULAR PARASITES PATHOGEN RESISTANCE INVOLVES SEQUESTRATION
CC OF FE2+ AND MN2+, COFACTORS OF BOTH PROKARYOTIC AND EUKARYOTIC
CC CATALASES AND SUPEROXIDE DISMUTASES, NOT ONLY TO PROTECT THE
CC MACROPHAGE AGAINST ITS OWN GENERATION OF REACTIVE OXYGEN SPECIES,
CC BUT TO DENY THE CATIONS TO THE PATHOGEN FOR SYNTHESIS OF ITS
CC PROTECTIVE ENZYMES (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Probable).
CC -1- SIMILARITY: BELONGS TO THE NRAMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U39614; AAB17552.1; -.
CC InterPro; IPR001046; Nrmamp.
CC Pfam; PF01566; Nrmamp; 1.
CC PRINTS; PR00447; NATRESASSCMP.
CC ProDom; PD001861; Nrmamp; 1.
CC Transprot; Iron transport; Transmembrane; Glycoprotein; Macrophage.
CC DOMAIN 1 55 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 56 73 POTENTIAL.
FT DOMAIN 74 82 EXTRACELLULAR (POTENTIAL).
```

```
FT TRANSMEM 83 102 POTENTIAL.
FT DOMAIN 103 139 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 140 160 POTENTIAL.
FT DOMAIN 161 164 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 165 184 POTENTIAL.
FT DOMAIN 185 193 POTENTIAL.
FT TRANSMEM 194 214 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 215 237 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 238 256 POTENTIAL.
FT DOMAIN 257 284 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 285 304 POTENTIAL.
FT DOMAIN 305 346 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 347 366 POTENTIAL.
FT DOMAIN 367 397 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 398 415 POTENTIAL.
FT DOMAIN 416 426 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 427 447 POTENTIAL.
FT DOMAIN 448 463 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 464 485 POTENTIAL.
FT DOMAIN 486 493 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 494 513 POTENTIAL.
FT DOMAIN 514 548 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 3 77 PRO/SER-RICH.
FT CARBOHYD 321 321 N-LINKED (GLCNAC... ) (POTENTIAL).
FT CARBOHYD 335 335 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 548 AA; 59529 MW; 33DF170E48D51878 CRC64;

Query Match 72.5%; Score 29; DB 1; Length 548;
Best Local Similarity 66.7%; Pred. No. 2.le+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
Db 433 SLLPLFAVL 441
|||||:|

Search completed: November 6, 2002, 12:09:15
Job time : 7.33333 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 11 seconds
(without alignments)
78.619 Million cell updates/sec

Title: US-09-905-083-36
Perfect score: 40
Sequence: 1 SLLPLQL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	253	2 A53968	serine proteinase
2	38	95.0	146	2 H75201	hypothetical prote
3	33	82.5	370	2 AB3334	daunorubicin resis
4	32	80.0	218	2 T02912	probable transcrip
5	32	80.0	264	2 C97402	probable acyltrans
6	32	80.0	264	2 AC2620	1-acyl-sn-glycerol
7	32	80.0	266	2 JC7300	tax-responsive ele
8	32	80.0	267	2 JC4857	hepatocarcinogenes
9	32	80.0	282	2 B85327	probable transcrip
10	32	80.0	494	2 C49349	probable succinogl
11	32	80.0	494	2 B95976	probable transport
12	32	80.0	582	2 S40176	ExoT protein - Rhi
13	32	80.0	622	2 S61692	probable membrane
14	31	77.5	186	2 C65039	hypothetical prote
15	31	77.5	303	2 C84914	hypothetical prote
16	31	77.5	354	2 T48649	glycerol-3-phospha
17	31	77.5	372	2 A98157	probable permease
18	31	77.5	372	2 AB3130	ABC transporter, m
19	31	77.5	392	2 D83934	hypothetical prote
20	31	77.5	395	2 D69779	antibiotic resista
21	31	77.5	398	2 C91063	hypothetical prote
22	31	77.5	401	2 B90120	SNF1-related prote
23	31	77.5	413	2 AC0834	probable membrane
24	31	77.5	470	2 A90083	hypothetical prote
25	31	77.5	475	1 A69149	O-antigen transpor
26	31	77.5	754	2 AE0614	probable competenc
27	31	77.5	783	2 A46136	myosin-heavy-chain
28	31	77.5	913	2 AC2445	hypothetical prote
29	30	75.0	155	2 S38662	interleukin-2 - go

30	30	75.0	155	2	S11488	interleukin-2 prec
31	30	75.0	155	2	I45913	interleukin-2 prec
32	30	75.0	168	2	E96979	probable membrane
33	30	75.0	189	2	T43766	hypothetical prote
34	30	75.0	190	2	T39622	probable phosphate
35	30	75.0	208	2	T17092	NADH dehydrogenase
36	30	75.0	214	1	G69798	hypothetical prote
37	30	75.0	223	2	F83598	cell division prot
38	30	75.0	277	2	H84314	cytochrome aa3 con
39	30	75.0	289	2	T43663	probable high affi
40	30	75.0	333	1	HLHUCB	T-cell surface gly
41	30	75.0	349	2	E82656	conserved hypotet
42	30	75.0	369	2	S33603	surfactant protein
43	30	75.0	371	1	JN0450	conglutinin precu
44	30	75.0	371	2	I45878	conglutinin - bovi
45	30	75.0	375	2	F91173	probable transport

ALIGNMENTS

RESULT 1

A53968
serine proteinase SCCE precursor - human
N:Alternate names: stratum corneum chymotryptic enzyme
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
C:Accession: A53968
R:Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud J. Biol. Chem. 269, 19420-19426, 1994
A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic e
A:Reference number: A53968; MUID:94308225
A:Accession: A53968
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253 <HAN>
A:Cross-references: GB:L33404; NID:g521214; PIDN:AAC37551.1; PID:g532504
C:Genetics:
A:Gene: GDB:PRSS6; SCCE
A:Cross-references: GDB:377730
A:Map position: 7q35-7q35
C:Superfamily: trypsin; trypsin homology
F:30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.6;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
|||||||
DB 4 SLLPLQL 12

RESULT 2

H75201
hypothetical protein PAB0088 - Pyrococcus abyssi (strain Orsay)
C:Species: Pyrococcus abyssi
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 08-Sep-2000
C:Accession: H75201
R:anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome
A:Reference number: A75001
A:Accession: H75201
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-146 <KAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:g5457433; PIDN:CAB49063.1; PID:el
C:Experimental source: strain Orsay
C:Genetics:
A:Gene: PAB0088
C:Superfamily: Pyrococcus abyssi hypothetical protein PAB0088

Query Match 95.0%; Score 38; DB 2; Length 146;
Best Local Similarity 88.9%; Pred. No. 2.2;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
 |||||:
Db 66 SLLPLQLII 74

RESULT 3

AB3334
daunorubicin resistance transmembrane protein [imported] - Brucella melitensis (strain 1
C:Species: Brucella melitensis
C>Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 01-Feb-2002
C:Accession: AB3334
R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
M.; Mazur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AB3334
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-370 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL51837.1; PID:gl7982584; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0656
A:Map position: 1

Query Match 82.5%; Score 33; DB 2; Length 370;
Best Local Similarity 75.0%; Pred. No. 57;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
 |||||:
Db 293 ILLPLQLV 300

RESULT 4

T02912
probable transcription factor MYB41 [similarity] - Arabidopsis thaliana (fragment)
N:Alternate names: protein T13J8.220
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Sep-2000
C:Accession: T02912; T51654
R:Bayan, M.; Pohl, T.; Weizenegger, T.; Hohnel, J.; Mewes, H.W.; Mayer, K.F.X.; Schuel
submitted to the Protein Sequence Database, February 1999
A:Reference number: Z14766
A:Accession: T02912
A:Molecule type: DNA
A:Residues: 1-218 <BEV>
A:Cross-references: EMBL:AL035524
A:Experimental source: cultivar Columbia; BAC clone T13J8
R:Kranz, H.D.; Denekamp, M.; Greco, R.; Jin, H.; Kranz, H.D.; Denekamp, M.; Greco, R.; J
Paz-Ares, J.; Weishaar, B.
Plant J. 16, 263-276, 1998
A:Title: Towards functional characterisation of the members of the R2R3-MYB gene from Ar
A:Reference number: Z14349; MUID:9839469
A:Accession: T51654
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-218 <KRA>
A:Cross-references: EMBL:AF062882; PIDN:AAC83604.1
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: MYB41
A:Map position: 4
A:Note: T13J8.220
A:Note: intron positions not resolved
C:Superfamily: unassigned myb DNA-binding repeat proteins; myb DNA-binding repeat homolo
C:Keywords: transcription factor
F:1-48/Domain: myb DNA-binding repeat homology (fragment) <MYB>

Query Match 80.0%; Score 32; DB 2; Length 218;
Best Local Similarity 100.0%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 7
 |||||:
Db 104 SLLPLQL 110

RESULT 5

C97402
probable acyltransferase (AF232919) [imported] - Agrobacterium tumefaciens (strain C5
C:Species: Agrobacterium tumefaciens
C>Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: C97402
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouello, B.; Goldm
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium
A:Reference number: A97359; PMID:11743194
A:Accession: C97402
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-264 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK86172.1; PID:gl5155265; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_621
A:Map position: circular chromosome

Query Match 80.0%; Score 32; DB 2; Length 264;
Best Local Similarity 75.0%; Pred. No. 64;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
 |||||:
Db 19 ILLPLQLL 26

RESULT 6

AC2620
1-acyl-sn-glycerol-3-phosphate acyltransferase plsc [imported] - Agrobacterium tumefa
C:Species: Agrobacterium tumefaciens
C>Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AC2620
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erger, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; MCCI
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-kam
ster, B.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AC2620
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-264 <KUR>
A:Cross-references: GB:AE008688; PIDN:AAL41377.1; PID:gl7738693; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: plsc
A:Map position: circular chromosome

Query Match 80.0%; Score 32; DB 2; Length 264;
Best Local Similarity 75.0%; Pred. No. 64;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
 |||||:
Db 19 ILLPLQLL 26

RESULT 7

JC7300
tax-responsive element-binding protein 5 - mouse
C:Species: Mus musculus (house mouse)
C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Sep-2000
C:Accession: JC7300
R:Masaki, T.; Noguchi, H.; Kobayashi, M.; Yoshida, M.; Takamatsu, K.
DNA Res. 7, 187-193, 2000
A>Title: Isolation and characterization of the gene encoding mouse tax-responsive element
A:Reference number: JC7300
A:Accession: JC7300
A:Molecule type: mRNA
A:Residues: 1-266 <MAS>
A:Cross-references: DDBJ:AB036745
C:Genetics:
A:Gene: treb5
C:Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology
C:Keywords: leucine zipper; transcription factor

Query Match 80.0%; Score 32; DB 2; Length 266;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQIL 9
|||||
Db 187 LLPLQIL 193

RESULT 8
JC4857
hepatocarcinogenesis-related transcription factor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 08-Sep-1997
C:Accession: JC4857
R:Kishimoto, T.; Kokura, K.; Kumagai, Y.; Makino, Y.; Tamura, T.
Biochem. Biophys. Res. Commun. 224, 746-751, 1996
A>Title: HTF: A b-zip transcription factor that is closely related to the human XBP/TRF
A:Reference number: JC4857
A:Accession: JC4857
A:Molecule type: mRNA
A:Residues: 1-267 <KIS>
C:Comment: This is a basic-leucine zipper type transcription factor involved in hepatocarcinogenesis
C:Genetics:
A:Gene: htf
C:Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology
C:Keywords: leucine zipper; transcription factor
F:58-98/Domain: fos/jun DNA-binding domain homology <FJD>
F:89-126/Region: leucine zipper motif

Query Match 80.0%; Score 32; DB 2; Length 267;
Best Local Similarity 100.0%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQIL 9
|||||
Db 188 LLPLQIL 194

RESULT 9
B85327
probable transcription factor MYB41 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 02-Mar-2001
C:Accession: B85327
R:anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Nature 402, 769-777, 1999
A>Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.
A:Reference number: A85001; MUID:20083488
A:Accession: B85327
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-282 <STO>
A:Cross-references: GB:NC_001268; NID:g7269665; PIDN:CAB79613.1; GSPDB:GN00140
C:Genetics:

A:Gene: AT4g28110
A:Map position: 4
C:Superfamily: barley myb-related protein 3; myb DNA-binding repeat homology

Query Match 80.0%; Score 32; DB 2; Length 282;
Best Local Similarity 100.0%; Pred. No. 68;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQ 7
|||||
Db 168 SLLPLQ 174

RESULT 10
C49349
probable succinoylcan transport protein ExoT - Rhizobium meliloti
C:Species: Rhizobium meliloti
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 17-Mar-2000
C:Accession: C49349
R:Glucksmann, M.A.; Reuber, T.L.; Walker, G.C.
J. Bacteriol. 175, 7045-7055, 1993
A>Title: Genes needed for the modification, polymerization, export, and processing of
A:Reference number: A49349; MUID:94042870
A:Accession: C49349
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <GLU>
A:Cross-references: GB:L20758; NID:g393240; PIDN:AAAL6050.1; PID:g393249
C:Superfamily: hypothetical protein b2046
C:Keywords: transmembrane protein

Query Match 80.0%; Score 32; DB 2; Length 494;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQ 8
|||||
Db 122 SLLPLQ 129

RESULT 11
B95976
probable transport protein, similar to Wzx exoT [imported] - Sinorhizobium meliloti
C:Species: Sinorhizobium meliloti
C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: B95976
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; He.
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A>Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: B95976
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-494 <KUR>
A:Cross-references: GB:AL591985; PIDN:CAC49474.1; PID:gl5140960; GSPDB:GN00167
A:Experimental source: Strain 1021, megaplasmid pSymb
R:Gallbert, F.; Finan, T.M.; Long, S.K.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hub
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lela
heault, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh
A>Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: exoT; Smb20950
A:Genome: plasmid
C:Superfamily: hypothetical protein b2046

Query Match 80.0%; Score 32; DB 2; Length 494;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQI 8
| | | | |
Db 122 SLLPLQL 129

RESULT 12
S40176
ExoT protein - Rhizobium meliloti
C:Species: Rhizobium meliloti
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 19-May-2000
C:Accession: S40176
R:Becker, A.; Kleickmann, A.; Kuester, H.; Keller, M.; Arnold, W.; Puehler, A.
submitted to the EMBL Data Library, April 1993
A:Description: Analysis of the Rhizobium meliloti genes exoU, exoV, exoW, exoT and exoI
ses.
A:Reference number: S40173
A:Accession: S40176
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-582 <BEC>
A:Cross-references: EMBL:Z22646
C:Superfamily: hypothetical protein b2046

Query Match 80.0%; Score 32; DB 2; Length 582;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQI 8
| | | | |
Db 122 SLLPLQL 129

RESULT 13
S61692
Probable membrane protein YOR137c - yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein O3329; hypothetical protein YOR3329c
C:Species: Saccharomyces cerevisiae
C:Date: 09-Mar-1996 #sequence_revision 12-Apr-1996 #text_change 21-Nov-1997
C:Accession: S61692; S67022
R:Benes, V.; Andrade, M.A.; Rechmann, S.; Teodoru, C.; Banrevi, A.; Sander, C.; Valencia
submitted to the EMBL Data Library, December 1995
A:Description: Nucleotide sequence and analysis of a 130 kb fragment of yeast chromosome
A:Reference number: S61643
A:Accession: S61692
A:Molecule type: DNA
A:Residues: 1-622 <BEN>
A:Cross-references: EMBL:X94335; NID:gl362139; PID:e217839; PID:gl164980
R:Voss, H.; Benes, V.; Rechmann, S.; Teodoru, C.; Schwager, C.; Paces, V.; Ansorge, W.
submitted to the Protein Sequence Database, July 1996
A:Reference number: S66965
A:Accession: S67022
A:Molecule type: DNA
A:Residues: 1-622 <VOS>
A:Cross-references: EMBL:Z75045; NID:gl420348; PID:e252028; PID:gl420349; MIPS:YOR137c
A:Experimental source: strain S288C
C:Genetics:
A:Map position: 15R
C:Keywords: transmembrane protein
F:11-27/Domain: transmembrane #status predicted <TMM>

Query Match 80.0%; Score 32; DB 2; Length 622;
Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
| | | | |
Db 141 SALIPLQL 149

RESULT 14
G65039
hypothetical protein b2612 - Escherichia coli (strain K-12)

C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: G65039
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G65039
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-196 <BLAT>
A:Cross-references: GB:AE000347; GB:U00096; NID:g2367142; PIDN:AACT5661.1; PID:g17889
A:Experimental source: strain K-12, substrain MG1655

Query Match 77.5%; Score 31; DB 2; Length 196;
Best Local Similarity 87.5%; Pred. No. 74;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
| | | | |
Db 123 LLAPLQL 130

RESULT 15
C84914
hypothetical protein At2g47360 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84914
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
A:Reference number: A84420; MUID:20083487
A:Accession: C84914
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-303 <STO>
A:Cross-references: GB:AE002093; NID:g2275214; PIDN:AAB63836.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g47360
A:Map position: 2

Query Match 77.5%; Score 31; DB 2; Length 303;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
| | | | |
Db 29 SLLPLSL 37

RESULT 16
T48649
glycerol-3-phosphate dehydrogenase (NAD+) (EC 1.1.1.8) [validated] - Trypanosoma brucei
C:Species: Trypanosoma brucei
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
C:Accession: T48649
R:Kohl, L.; Drmota, T.; Thi, C.D.; Callens, M.; Van Beeumen, J.; Oppendoes, F.R.; Mic
Mol. Biochem. Parasitol. 76, 159-173, 1996
A:Title: Cloning and characterization of the NAD-linked glycerol-3-phosphate dehydrog
scherichia coli.
A:Reference number: Z24504; MUID:97077437
A:Accession: T48649
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-354 <KOH>
A:Cross-references: EMBL:X89738; PIDN:CAA61890.1
A:Experimental source: ssp. brucei; strain 427
C:Genetics:
A:Gene: gpd

C:Function:

A:Description: EC 1.1.1.1.8 [validated, MUID:97077437]

C:Keywords: oxidoreductase

Query Match 77.5%; Score 31; DB 2; Length 354;

Best Local Similarity 77.8%; Pred. No. 1.4e+02;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

|||||

Db 124 SLLFPAQL 132

RESULT 17

A98157

probable permease of ABC transporter PA5230 [imported] - Agrobacterium tumefaciens (stra

C:Species: Agrobacterium tumefaciens

C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002

C:Accession: A98157

R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ourullo, B.; Goldman,

Science 294, 2323-2328, 2001

A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194

A:Accession: A98157

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-372 <KUR>

A:Cross-references: GB:AE007870; PIDN:AAK88779.1; PID:g15158527; GSPDB:GN00170

C:Genetics:

A:Gene: AGR_L 429

A:Map position: linear chromosome

Query Match 77.5%; Score 31; DB 2; Length 372;

Best Local Similarity 85.7%; Pred. No. 1.4e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQL 9

|||||

Db 296 LLPLQL 302

RESULT 18

AH3130

ABC transporter, membrane spanning protein Atu4668 [imported] - Agrobacterium tumefaciens

C:Species: Agrobacterium tumefaciens

C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002

C:Accession: AH3130

R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L

erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell

; Karp, P.; Romero, P.; Zhang, S.

Science 294, 2317-2323, 2001

A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,

ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193

A:Accession: AH3130

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-372 <KUR>

A:Cross-references: GB:AE008689; PIDN:AAL45462.1; PID:g17743167; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: Atu4668

A:Map position: linear chromosome

Query Match 77.5%; Score 31; DB 2; Length 372;

Best Local Similarity 85.7%; Pred. No. 1.4e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQL 9

|||||

Db 296 LLPLQL 302

RESULT 19

D83934

hypothetical protein BH2276 [imported] - Bacillus halodurans (strain C-125)

C:Species: Bacillus halodurans

C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: D83934

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: D83934

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-332 <STO>

A:Cross-references: GB:AP001515; GB:BA000004; NID:gl0174886; PIDN:BAB05995.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH2276

Query Match 77.5%; Score 31; DB 2; Length 392;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQL 9

|||||

Db 290 LLLPLQL 297

RESULT 20

D69779

antibiotic resistance protein homolog ydeR - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000

C:Accession: D69779

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be

C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M

Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinc

A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scani

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiya

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtili

A:Reference number: A69580; MUID:98044033

A:Accession: D69779

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-395 <KUN>

A:Cross-references: GB:Z99106; GB:AL009126; NID:g2632653; PIDN:CAB12338.1; PID:g26328

A:Experimental source: strain 168

C:Genetics:

A:Gene: ydeR

C:Superfamily: probable antibiotic resistance protein yybF

Query Match 77.5%; Score 31; DB 2; Length 395;

Best Local Similarity 87.5%; Pred. No. 1.5e+02;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 8

|||||

Db 294 SLLPLQL 301

RESULT 21

C91063

```

hypothetical protein ECs3475 [imported] - Escherichia coli (strain O157:H7, substrain R)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 24-Aug-2001
C:Accession: C91063
R:Hayaashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: C91063
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-398 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA036898.1; PID:gl3362946; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RIMD 0509952
C:Genetics:
A:Gene: ECs3475
C:Superfamily: hypothetical protein HI0107

Query Match 77.5%; Score 31; DB 2; Length 398;
Best Local Similarity 87.5%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
   || |||||
Db 101 LLAPLQL 108

RESULT 22
B90120
SNF1-related protein kinase [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
C:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
A:Title: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: B90120
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.; Re
Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; MUID:11323671
A:Accession: B90120
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-401 <DOU>
A:Cross-references: GB:AF083031; NID:gl3794278; PIDN:AAK39655.1; GSPDB:GN00152
C:Genetics:
A:Gene: kin(snf2)
A:Map position: 3
A:Genome: nucleomorph
C:Keywords: nucleomorph

Query Match 77.5%; Score 31; DB 2; Length 401;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
   | |||||
Db 304 SFLPLQL 312

RESULT 23
AC0834
probable membrane protein corB [imported] - Salmonella enterica subsp. enterica serovar
C:Species: Salmonella enterica subsp. enterica serovar typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AC0834
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
th, T.; Connor, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
, S.; Moule, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quall, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608

```

```

A:Accession: AC0834
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-413 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05858.1; PID:g16503833; GSPDB:GN00176
C:Genetics:
A:Gene: corB
C:Superfamily: hypothetical protein HI0107

```

```

Query Match 77.5%; Score 31; DB 2; Length 413;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

QY 2 LLLPLQL 9
   || |||||
Db 116 LLAPLQL 123

```

RESULT 24

```

A90083
hypothetical protein orf470 [imported] - Guillardia theta nucleomorph
C:Species: nucleomorph Guillardia theta
A:Note: a nucleomorph is the vestigial nucleus of a eukaryotic endosymbiont
C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 24-May-2001
C:Accession: A90083
R:Douglas, S.; Zauner, S.; Fraunholz, M.; Beaton, M.; Penny, S.; Deng, L.T.; Wu, X.;
Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.
A:Reference number: A99082; MUID:11323671
A:Accession: A90083
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-470 <DOU>
A:Cross-references: GB:AF165818; NID:g6690603; PIDN:AAF24211.1; GSPDB:GN00150
C:Genetics:
A:Gene: orf470
A:Map position: 1
A:Genome: nucleomorph
C:Keywords: nucleomorph

```

```

Query Match 77.5%; Score 31; DB 2; Length 470;
Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 2 LLLPLQL 9
   || |||||
Db 253 MLLPLEIL 260

```

RESULT 25

```

A69149
O:antigen transporter related protein - Methanobacterium thermoautotrophicum (strain
C:Species: Methanobacterium thermoautotrophicum
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: A69149
R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.
; Qiu, D.; Spadafora, R.; Vicaire, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jilwani,
ki, S.; Church, G.W.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
J. Bacteriol. 179, 7135-7155, 1997
A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: fu
A:Reference number: A69000; MUID:98037514
A:Accession: A69149
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-475 <MTH>
A:Cross-references: GB:AE000823; GB:AE000666; NID:g2621432; PIDN:AAB84885.1; PID:g262
A:Experimental source: strain Delta H
C:Genetics:
A:Gene: MTH379
A:Start codon: GTG
C:Superfamily: succinoglycan biosynthesis transport protein

```

```

Query Match 77.5%; Score 31; DB 1; Length 475;

```

Best Local Similarity 87.5%; Pred. No. 1.9e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
||||| |

Db 432 LLLPLSL 439

RESULT 26
AE0614
probable competence-related protein STY0984 [imported] - Salmonella enterica subsp. ente
C:Species: Salmonella enterica subsp. enterica serovar Typhi
A:Note: this species has also been called Salmonella typhi
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 09-Nov-2001
C:Accession: AE0614
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
Th, F.; Conerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Moulé, S.; O'Gaora, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A:Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A:Reference number: AB0502; PMID:11677608
A:Accession: AE0614
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-754 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD05383.1; PID:g16502146; GSPDB:GN00176
C:Genetics:
A:Gene: STY0984

Query Match 77.5%; Score 31; DB 2; Length 754;
Best Local Similarity 85.7%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQI 8
||||| |

Db 371 LLLPLQV 377

RESULT 27
A46136
myosin-heavy-chain kinase (EC 2.7.1.129) - slime mold (Dictyostelium discoideum)
C:Species: Dictyostelium discoideum
C:Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 17-Apr-1998
C:Accession: A46136
R:Ravid, S.; Spudich, J.A.
Proc. Natl. Acad. Sci. U.S.A. 89, 5877-5881, 1992
A:Title: Membrane-bound Dictyostelium myosin heavy chain kinase: a developmentally regul
A:Reference number: A46136; MUID:92335200
A:Contents: AX3
A:Accession: A46136
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-783 <RAV>
A:Cross-references: GB:M93393; NID:g167827
A:Note: sequence extracted from NCBI backbone (NCBIN:108552, NCBI:P:108553); this ORF is
C:Superfamily: protein kinase C zinc-binding repeat homology
C:Keywords: phosphotransferase
F:8-138/Domain: protein kinase C zinc-binding repeat homology <K22>

Query Match 77.5%; Score 31; DB 2; Length 783;
Best Local Similarity 77.8%; Pred. No. 3.1e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
||| ||||

Db 236 SLLNPLQII 244

RESULT 28
AC2445
hypothetical protein all5115 [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC2445
R:Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irig
Nakazaki, N.; Shimpō, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabat
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2445
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-913 <KUR>
A:Cross-references: GB:BA000019; PIDN:BA076814.1; PID:g17134253; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: all5115

Query Match 77.5%; Score 31; DB 2; Length 913;
Best Local Similarity 77.8%; Pred. No. 3.6e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
||||| |

Db 18 SLLPLGLII 26

RESULT 29
S38662
interleukin-2 - goat
C:Species: Capra aegagrus hircus (domestic goat)
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 16-Jul-1999
C:Accession: S38662
R:Rimstad, E.
submitted to the EMBL Data Library, November 1993
A:Description: The molecular cloning and expression of caprine interleukin 2.
A:Reference number: S38662
A:Accession: S38662
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-155 <RIM>
A:Cross-references: EMBL:X76063; NID:g416002; PIDN:CAA53664.1; PID:g416003
C:Superfamily: interleukin-2

Query Match 75.0%; Score 30; DB 2; Length 155;
Best Local Similarity 77.8%; Pred. No. 92;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
||||| |

Db 36 SLLDLQLL 44

RESULT 30
S11488
interleukin-2 precursor - sheep
C:Species: Ovis orientalis aries, Ovis ammon aries (domestic sheep)
C:Date: 21-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 16-Jul-1999
C:Accession: S11488; S13102; S15517
R:Goodall, J.C.; Emery, D.C.; Perry, A.C.F.; English, L.S.; Hall, L.
Nucleic Acids Res. 18, 5883, 1990
A:Title: cDNA cloning of ovine interleukin 2 by PCR.
A:Reference number: S11488; MUID:91016933
A:Accession: S11488
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-155 <GOO>
A:Cross-references: EMBL:X53934; NID:g1281; PIDN:CAA37881.1; PID:g1282
R:Scow, H.F.; Rothel, J.S.; Radford, A.J.; Wood, P.R.
Nucleic Acids Res. 18, 7175, 1990
A:Title: The molecular cloning of ovine interleukin 2 gene by the polymerase chain r
A:Reference number: S13102; MUID:91088336
A:Accession: S13102
A:Status: preliminary

A:Molecule type: mRNA
A:Residues: 1-5, 'L', 7-155 <SEO>
A:Cross-references: EMBL:X55641; NID:gl1810; PIDN:CAA39165.1; PID:g1811
R:Bujdosó, R.; Williamson, M.L.; Sargan, D.R.; Hein, W.H.; McConnell, I.
submitted to the EMBL Data Library, April 1991

A:Reference number: S15517
A:Accession: S15517
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 21-153 <BUJ>
A:Cross-references: EMBL:X60148
C:Superfamily: Interleukin-2

Query Match 75.0%; Score 30; DB 2; Length 155;
Best Local Similarity 77.8%; Pred. No. 92;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
||| |||
Db 36 SLLDLQL 44

RESULT 31

I45913
Interleukin-2 precursor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 16-Jul-1999
C:Accession: I45913; S21470; S20761
R:Corratti, D.P.; McKerghan, K.; Larsen, A.; Cantrell, M.A.; Anderson, D.; Gillis, S.;
Proc. Natl. Acad. Sci. U.S.A. 83, 3223-3227, 1986
A:Title: Cloning, sequence, and expression of bovine interleukin 2.
A:Reference number: I45913; MUID:86205869

A:Accession: I45913
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: mRNA
A:Residues: 1-155 <CER>
A:Cross-references: GB:M12791; NID:gl63204; PIDN:AAA30586.1; PID:g163205
R:Anikeeva, N.N.; Vinogradova, T.V.; Votoshin, O.N.
submitted to the EMBL Data Library, December 1989

A:Reference number: S21470
A:Accession: S21470
A:Molecule type: DNA
A:Residues: 1-22 <AN2>
A:Cross-references: EMBL:X17201; NID:g452; PIDN:CAA35062.1; PID:g453
C:Genetics:
A:Gene: IL-2
C:Superfamily: Interleukin-2

C:Keywords: cytokine; glycoprotein; growth factor; immunoregulation; lymphokine; T-cell

Query Match 75.0%; Score 30; DB 2; Length 155;
Best Local Similarity 77.8%; Pred. No. 92;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
||| |||
Db 36 SLLDLQL 44

RESULT 32

E96979
probable membrane protein [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: E96979
R:Nolling, J.; Breton, G.; Omeichenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,
; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: E96979
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-168 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK78624.1; PID:gl5023521; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC0647

Query Match 75.0%; Score 30; DB 2; Length 168;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
||| |||
Db 39 SLVPLSL 47

RESULT 33

T43766
hypothetical protein 189 [imported] - slime mold (Dictyostellum discoideum) mitochond
C:Species: Dictyostellum discoideum
C:Date: 31-Jan-2000 #sequence_revision 31-Jan-2000 #text_change 19-May-2000
C:Accession: T43766

R:Ogawa, S.; Yoshino, R.; Angata, K.; Pi, M.; Iwamoto, M.; Kuroe, K.; Matsuo, K.; Mor
submitted to the EMBL Data Library, December 1996
A:Description: The mitochondrial DNA of Dictyostellum discoideum. Complete sequence,
A:Reference number: Z22666

A:Accession: T43766
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-189 <OGA>
A:Cross-references: EMBL:AB000109; PIDN:BA78070.1

C:Genetics:
A:Superfamily: Dictyostellum mitochondrion hypothetical protein 189
C:Keywords: mitochondrion

Query Match 75.0%; Score 30; DB 2; Length 189;
Best Local Similarity 85.7%; Pred. No. 1.e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LLPLQL 9
:|||||
Db 106 ILPLQL 112

RESULT 34

T39622
probable phosphate transport protein - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 21-Jan-2000
C:Accession: T39622

R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL Data Library, March 1998

A:Reference number: Z21843

A:Accession: T39622

A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-190 <WOO>

A:Cross-references: EMBL:AL022104; PIDN:CAA17902.1; GSPDB:GN00067; SPDB:SPBC16H5.04

A:Experimental source: strain 972h-; cosmid c16H5

C:Genetics:

A:Gene: SPDB:SPBC16H5.04

A:Map position: 2

A:Introns: 13/3; 47/3

C:Superfamily: Saccharomyces hypothetical protein YBR106W

Query Match 75.0%; Score 30; DB 2; Length 190;
Best Local Similarity 77.8%; Pred. No. 1.e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
||| |||
Db 36 SLLPLRL 44

RESULT 35
Tl7092
C:Species: NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) chain 4 - Sigmodon hispidus mitochondrion
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 26-May-2000
C:Accession: Tl7092
R:Engel, S.R.; Hogan, K.M.; Taylor, J.F.; Davis, S.K.
submitted to the EMBL Data Library, January 1997
A:Description: Molecular systematics and paleobiogeography of the South American sigmoid
A:Reference number: Z18675
A:Accession: Tl7092
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-208 <ENG>
A:Cross-references: EMBL:U03823; NID:g2645362; PID:g2645365; PIDN:AAB87240.1
C:Genetics:
A:Genome: mitochondrion
A:Note: ND4
C:Keywords: membrane-associated complex; mitochondrion; NAD; oxidative phosphorylation;
Query Match 75.0% Score 30; DB 2; Length 208;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 SLLPLQL 9
::: |||||
Db 97 SLLISLQL 105
RESULT 36
G69798
hypothetical protein yetJ - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: G69798
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertsch
C.; Bron, S.D.; Bouilliet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Ch
A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabet, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gall
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
Koetter, P.; Koningsstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois,
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maue
Y, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetel
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon,
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seron
akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis.
A:Reference number: A69580; MUID:98044033
A:Accession: G69798
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-214 <KUN>
A:Cross-references: GB:AL009126; NID:g2632866; PIDN:CAB12539.1; PID:g2633033
A:Experimental source: Strain 168
C:Genetics:
A:Gene: yetJ
C:Superfamily: hypothetical protein yetJ
Query Match 75.0% Score 30; DB 1; Length 214;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 SLLPLQL 9
::: |||||
Db 40 ALMLPLSL 48
RESULT 37
F83598
cell division protein FtsE PA0374 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83598
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p
A:Reference number: A82950; MUID:20437337
A:Accession: F83598
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-223 <STO>
A:Cross-references: GB:AE004475; GB:AE004091; NID:g9946221; PIDN:AAG03763.1; GSPDB:J
A:Experimental source: strain PA01
C:Genetics:
A:Gene: ftsE; PA0374
Query Match 75.0% Score 30; DB 2; Length 223;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 SLLPLQL 9
::: |||||
Db 99 NIALPLQL 107
RESULT 38
H84314
cytochrome aa3 controlling protein [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84314
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lask
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; J
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.
A:Title: Genome sequence of Halobacterium species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: H84314
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-277 <STO>
A:Cross-references: GB:AE004437; NID:q10581096; PIDN:AAG19884.1; GSPDB:GN00138
C:Genetics:
A:Gene: ccp
Query Match 75.0% Score 30; DB 2; Length 277;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
Qy 2 LLLPLQL 9
::: |||||
Db 97 VLLPLQVI 104
RESULT 39
T43663
probable high affinity copper transport protein ctr4 - fission yeast (Schizosaccharo
C:Species: Schizosaccharomyces pombe
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000
C:Accession: T43663; T40958; T51300
R:Labbe, S.; Pena, M.M.; Fernandes, A.R.; Thiele, D.J.
submitted to the EMBL Data Library, August 1999
A:Description: A copper sensing transcription factor regulates iron uptake genes in
A:Reference number: Z22610
A:Accession: T43663
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-289 <LAB>
A:Cross-references: EMBL:AJ243833; PIDN:CAB52305.1

R;Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
submitted to the EMBL Data Library, February 1999
A;Reference number: 221940
A;Accession: T40958
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-289 <WO>
A;Cross-references: EMBL:AJ035592; PIDN:CAB38165.1; GSPDB:GN00068; SPDB:SPCC1393.10
A;Experimental source: strain 972h-; cosmid c1393
R;Labbe, S.; Pena, M.M.O.; Fernandes, A.R.; Thiele, D.J.
submitted to the EMBL Data Library, August 1999
A;Description: A copper sensing transcription factor regulates iron uptake genes in *Schistosoma mansoni*
A;Reference number: 225363
A;Accession: T51300
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-289 <LA2>
A;Cross-references: EMBL:AF175405; PIDN:AAD51064.1
A;Experimental source: strain FY254
C;Genetics:
A;Gene: ctr4; SPCC1393.10
A;Map position: 3

Query Match 75.0%; Score 30; DB 2; Length 289;
Best Local Similarity 77.8%; Pred. No. 1.8e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
||| |||||
Db 40 NLLPLQL 48

RESULT 40

HLHUCB
T-cell surface glycoprotein CD1b precursor - human
N;Alternate names: thymocyte differentiation antigen CD1b
C;Species: Homo sapiens (man)
C;Date: 06-Mar-1992 #sequence_revision 17-May-1996 #text_change 22-Jun-1999
C;Accession: B39957; PIDN:179470
R;Martin, L.H.; Calabi, F.; Lefebvre, F.A.; Bilsland, C.A.G.; Milstein, C.
Proc. Natl. Acad. Sci. U.S.A. 84, 9189-9193, 1987
A;Title: Structure and expression of the human thymocyte antigens CD1a, CD1b, and CD1c.
A;Reference number: A39957; MUID:88097453
A;Accession: B39957
A;Molecule type: DNA
A;Residues: 1-333 <MAR>
A;Cross-references: GB:M22173; GB:J03584; NID:g180062; PIDN:AAA51940.1; PID:g180064
A;Experimental source: Clone lambda RIB5
R;Aruffo, A.; Seed, B.
J. Immunol. 143, 1723-1730, 1989
A;Title: Expression of cDNA clones encoding the thymocyte antigens CD1a, b, c demonstrated in thymocytes
A;Reference number: A45801; MUID:89341413
A;Accession: B45801
A;Molecule type: mRNA
A;Residues: 1-333 <ARU>
A;Cross-references: GB:M28826; NID:g180055; PIDN:AAA51939.1; PID:g180056
R;Martin, L.H.; Calabi, F.; Milstein, C.
Proc. Natl. Acad. Sci. U.S.A. 83, 9154-9158, 1986
A;Title: Isolation of CD1 genes: a family of major histocompatibility complex-related genes
A;Reference number: I59088; MUID:87067468
A;Accession: I79470
A;Status: translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 203-295 <RES>
A;Cross-references: GB:M14665; NID:g180049; PIDN:AAA51936.1; PID:g180050
C;Genetics:
A;Gene: GDB:CD1B; CDI
A;Cross-references: GDB:I20576; OMIM:188360
A;Map position: 1q21-1q23
A;Introns: 21/1; 110/1; 203/1; 296/1; 327/2
C;Complex: heterodimer with beta-2-microglobulin (see PIR:MGHUB2)
C;Superfamily: class I histocompatibility antigen; immunoglobulin homology
C;Keywords: glycoprotein; heterodimer; thymocyte; transmembrane protein

F;1-18/Domain: signal sequence #status predicted <SIG>
F;19-333/Product: T-cell surface glycoprotein CD1b #status predicted <MAY>
F;19-109/Domain: alpha-1 <EX1>
F;110-202/Domain: alpha-2 <EX2>
F;203-295/Domain: alpha-3 <EX3>
F;216-281/Domain: immunoglobulin homology <IMM>
F;301-324/Domain: transmembrane #status predicted <TRM>
F;325-333/Domain: intracellular #status predicted <INT>
F;38,75,146,258/Binding site: carbohydrate (Asn) (covalent) #status predicted
F;120-184,224-279/Disulfide bonds: #status predicted

Query Match 75.0%; Score 30; DB 1; Length 333;
Best Local Similarity 75.0%; Pred. No. 2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
||||| |||
Db 2 LLLPLQL 9

RESULT 41

E82656
conserved hypothetical protein XF1630 [Imported] - *Xylella fastidiosa* (strain 9a5c)
C;Species: *Xylella fastidiosa*
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000
C;Accession: E82656
R;anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing
Nature 406, 151-157, 2000
A;Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.
A;Reference number: A82515; MUID:20365717
A;Note: for a complete list of authors see reference number A59328 below
A;Accession: E82656
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-349 <SIM>
A;Cross-references: GB:AE003990; GB:AE003849; NID:g9106683; PIDN:AAF84439.1; GSPDB:GN
R;Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer
as-Neto, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to Genbank, June 2000
A;Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A;Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A;Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A;Reference number: A59328
A;Contents: annotation
C;Genetics:
A;Gene: XF1630

Query Match 75.0%; Score 30; DB 2; Length 349;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 7
|||||
Db 115 SLLPLQL 121

RESULT 42

S33603
surfactant protein D - bovine
C;Species: *Bos primigenius taurus* (cattle)
C;Date: 02-Dec-1993 #sequence_revision 01-Sep-1995 #text_change 17-Mar-1999
C;Accession: S33603
R;Lim, B.L.; Lu, J.; Reid, K.B.M.
Immunology 78, 159-165, 1993
A;Title: Structural similarity between bovine conglutinin and bovine lung surfactant
A;Reference number: S33603; MUID:93170856

A:Accession: S33603

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-369 <LIM>

C:Superfamily: pulmonary surfactant protein D; C-type lectin homology

F:248-367/Domain: C-type lectin homology <LCH>

Query Match 75.0%; Score 30; DB 2; Length 369;

Best Local Similarity 75.0%; Pred. No. 2.3e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLPLPQL 9

||||| :|

Db 2 LLLPLSVL 9

RESULT 43

JN0450

conglutinin precursor - bovine

N:Alternate names: C3b-binding protein

C:Contains: conglutinin-N

C:Species: Bos primigenius taurus (cattle)

C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 16-Jun-2000

C:Accession: JN0450; JC2396; S33235; A23740; S36879; S35044; I46010; A29416; S34054

R:Suzuki, Y.; Yin, Y.P.; Makino, M.; Kurimura, T.; Wakamiya, N.

Biochem. Biophys. Res. Commun. 191, 335-342, 1993

A:Title: Cloning and sequencing of a cDNA coding for bovine conglutinin.

A:Reference number: JN0450; MUID:93213261

A:Accession: JN0450

A:Molecule type: mRNA

A:Residues: 1-371 <SUZ>

A:Cross-references: DDBJ:D14085; NID:g285643; PIDN:BAA03170.1; PID:g285644

A:Experimental source: liver

R:Kawasaki, N.; Itoh, N.; Kawasaki, T.

Biochem. Biophys. Res. Commun. 198, 597-604, 1994

A:Title: Gene organization and 5'-flanking region sequence of conglutinin: A C-type mann

A:Reference number: JC2396; MUID:94128104

A:Accession: JC2396

A:Molecule type: mRNA

A:Residues: 1-371 <KA2>

A>Note: The authors translated the codon GAT for residues 250 and 270 as Glu

R:Lu, J.; Laursen, S.B.; Thiel, S.; Jensenius, J.C.; Reid, K.B.M.

Biochem. J. 292, 157-162, 1993

A:Title: The cDNA cloning of conglutinin and identification of liver as a primary site o

A:Reference number: S33235; MUID:93277452

A:Accession: S33235

A:Molecule type: mRNA

A:Residues: 1-172,'H',174-217,'A',219-271,'V',273-371 <LUJ>

A:Cross-references: EMBL:X71774; NID:g395267; PIDN:CAA50665.1; PID:g395268

A:Experimental source: liver

R:Lee, Y.M.; Leiby, K.R.; Allar, J.; Paris, K.; Lerch, B.; Okarma, T.B.

J. Biol. Chem. 266, 2715-2723, 1991

A:Title: Primary structure of bovine conglutinin, a member of the C-type animal lectin f

A:Reference number: A23740; MUID:91131556

A:Accession: A23740

A:Molecule type: protein

A:Residues: 21-209,'S',211-371 <LEE>

R:Kawasaki, N.; Yokota, Y.; Kawasaki, T.

Arch. Biochem. Biophys. 305, 533-540, 1993

A:Title: Differentiation of conglutination activity and sugar-binding activity of conglu

A:Reference number: S36879; MUID:93384312

A:Accession: S36879

A:Molecule type: protein

A:Residues: 21-54;75-86,'X',88-89,'X',91,'X',93-94;208-209,'X',211-227 <KAW>

A:Experimental source: serum

R:Lu, J.; Wiedemann, H.; Holmskov, U.; Thiel, S.; Timpl, R.; Reid, K.B.M.

Eur. J. Biochem. 215, 793-799, 1993

A:Title: Structural similarity between lung surfactant protein D and conglutinin. Two di

A:Reference number: S35044; MUID:93358905

A:Accession: S35044

A:Molecule type: protein

A:Residues: 75-86,'X',88-89,'X',91,'I' <LUA>

A:Experimental source: lung

R:Young, N.M.; Leon, M.A.

Biochem. Biophys. Res. Commun. 143, 645-651, 1987

A:Title: The carbohydrate specificity of conglutinin and its homology to proteins in

A:Reference number: A29416; MUID:87184551

A:Contents: annotation

R:Malhotra, R.; Laursen, S.B.; Willis, A.C.; Sim, R.B.

Biochem. J. 293, 15-19, 1993

A:Title: Research Communication. Localization of the receptor-binding site in the co

A:Reference number: S34054; MUID:93319501

A:Contents: annotation

R:Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sa.

J. Immunol. 153, 173-180, 1994

A:Title: Bovine conglutinin gene exon structure reveals its evolutionary relationship

A:Reference number: I46010; MUID:94267222

A:Accession: I46010

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-371 <LIO>

A:Cross-references: EMBL:U06860; NID:g507183; PIDN:AAB60624.1; PID:g514256

C:Comment: This protein mediates the agglutination of erythrocytes with antibody and

C:Comment: This protein is a Ca2+-dependent serum lectin specific for N-acetylglucos

C:Genetics:

A:Gene: CGN1

A:Introns: 67/1; 106/1; 142/1; 178/1; 217/1; 245/1

C:Superfamily: pulmonary surfactant protein D; C-type lectin homology

C:Keywords: calcium binding; glycoprotein; hydroxylysine; hydroxyproline

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-371/Product: conglutinin #status predicted <MAT>

F:46-214/Region: collagen-like

F:75-371/Product: conglutinin-N #status predicted <MA2>

F:248-369/Domain: C-type lectin homology <LCH>

F:63,87,99,135,141,159,162,198,210/Binding site: carbohydrate (Lys) (covalent) #statu

F:63,87,99,135,141,159,162,198,210/Modified site: 5-hydroxylysine (Lys) #status exper

F:78,96,108,111,129,132,147,153,171,195/Modified site: 4-hydroxyproline (pro) #statu

F:337/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 75.0%; Score 30; DB 1; Length 371;

Best Local Similarity 75.0%; Pred. No. 2.3e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLPLPQL 9

||||| :|

Db 2 LLLPLSVL 9

RESULT 44

I45878

conglutinin - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 20-Aug-1999

C:Accession: I45878

R:Liou, L.S.; Sastry, R.; Hartshorn, K.L.; Lee, Y.M.; Okarma, T.B.; Tauber, A.I.; Sa.

Gene 141, 277-281, 1994

A:Title: Bovine conglutinin (BC) mRNA expressed in liver: cloning and characterizati

A:Reference number: I45878; MUID:94215917

A:Accession: I45878

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-371 <LIO>

A:Cross-references: GB:I18871; NID:g495012; PIDN:AAA20126.1; PID:g495013

C:Superfamily: pulmonary surfactant protein D; C-type lectin homology

F:248-369/Domain: C-type lectin homology <LCH>

Query Match 75.0%; Score 30; DB 2; Length 371;

Best Local Similarity 75.0%; Pred. No. 2.3e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLPLPQL 9

||||| :|

Db 2 LLLPLSVL 9

RESULT 45

F91173
 probable transporter ECs4358 [imported] - Escherichia coli (strain O157:H7, substrain R
 C;Species: Escherichia coli
 C;Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C;Accession: F91173
 R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 DNA Res. 8, 11-22, 2001
 A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
 A;Reference number: A99629; MUID:21156231; PMID:11258796
 A;Accession: F91173
 A;Status: preliminary
 A;Molecule type: DNA
 A;Residues: 1-375 <HAY>
 A;Cross-references: GB:BA000007; PIDN:BA037781.1; PID:q13363832; GSPDB:GN00154
 A;Experimental source: strain O157:H7, substrain RMD 0509952
 C;Genetics:
 A;Gene: ECs4358

Query Match 75.0%; Score 30; DB 2; Length 375;
 Best Local Similarity 75.0%; Pred. No. 2.3e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLPLQL 9
 :|||||:
 Db 295 VLLPLQML 302

Search completed: November 6, 2002, 12:08:18
 Job time : 13 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 8.11111 Seconds
(without alignments)
27.102 Million cell updates/sec

Title: US-09-905-083-36
Perfect score: 40
Sequence: 1 SLLPLQL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40	100.0	9	4	US-09-502-600-36
2	40	100.0	253	2	US-08-557-146-2
3	40	100.0	253	2	US-08-824-874-3
4	40	100.0	253	2	US-09-154-344-2
5	40	100.0	253	3	US-08-930-188-2
6	40	100.0	253	4	US-09-210-084-3
7	40	100.0	253	5	PCT-US96-04294-2
8	36	90.0	9	4	US-09-502-600-33
9	32	80.0	9	4	US-09-502-600-35
10	32	80.0	9	4	US-09-502-600-116
11	31	77.5	307	4	US-09-177-249-13
12	30	75.0	135	2	US-08-383-621-5
13	30	75.0	135	3	US-08-459-906-5
14	30	75.0	463	2	US-08-677-049-6
15	30	75.0	549	2	US-08-676-279-57
16	30	75.0	1098	1	US-07-777-715-7
17	30	75.0	1098	1	US-08-170-126-2
18	30	75.0	1098	3	US-08-954-418-2
19	29	72.5	21	3	US-08-753-007A-24
20	29	72.5	21	4	US-09-398-496-24
21	29	72.5	22	3	US-08-977-378-22
22	29	72.5	26	3	US-08-630-172-40
23	29	72.5	26	4	US-09-375-419-40
24	29	72.5	190	1	US-08-339-152A-19
25	29	72.5	190	2	US-08-007-999B-6
26	29	72.5	190	2	US-08-689-276A-6
27	29	72.5	235	1	US-07-940-605A-12

28	72.5	235	2	US-08-690-096-12	Sequence 12, Appl
29	72.5	356	4	US-08-765-907A-8	Sequence 8, Appl
30	72.5	376	3	US-08-751-512-8	Sequence 8, Appl
31	72.5	419	4	US-09-554-225-1	Sequence 1, Appl
32	72.5	483	4	US-08-637-823B-4	Sequence 4, Appl
33	72.5	484	4	US-08-637-823B-2	Sequence 2, Appl
34	72.5	547	3	US-08-903-139B-7	Sequence 7, Appl
35	72.5	548	2	US-08-676-279-50	Sequence 50, Appl
36	72.5	548	3	US-08-903-139B-8	Sequence 8, Appl
37	72.5	548	3	US-08-903-139B-9	Sequence 9, Appl
38	72.5	548	3	US-08-903-139B-28	Sequence 28, Appl
39	72.5	548	4	US-08-637-823B-25	Sequence 25, Appl
40	72.5	550	4	US-08-637-823B-30	Sequence 30, Appl
41	72.5	634	1	US-08-339-152A-17	Sequence 17, Appl
42	72.5	653	1	US-08-339-152A-16	Sequence 16, Appl
43	72.5	653	2	US-08-007-999B-3	Sequence 3, Appl
44	72.5	653	2	US-08-689-276A-3	Sequence 3, Appl
45	72.5	4302	3	US-08-658-136-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-502-600-36
: Sequence 36, Application US/09502600A
: Patent No. 6294344
: GENERAL INFORMATION:
: APPLICANT: O'Brien, Timothy J.
: TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
: TITLE OF INVENTION: Ovarian Cancer
: FILE REFERENCE: D6223CIP-C
: CURRENT FILING DATE: 2000-02-11
: CURRENT APPLICATION NUMBER: US/09/502,600A
: PRIOR APPLICATION NUMBER: 09/039,211
: PRIOR FILING DATE: 03-14-1998
: NUMBER OF SEQ ID NOS: 136
: SEQ ID NO 36
: LENGTH: 9
: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: Residues 4-12 of the SCCE protein
US-09-502-600-36

Query Match 100.0%; Score 40; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
Db 1 SLLPLQL 9

RESULT 2
US-08-557-146-2
: Sequence 2, Application US/08557146
: Patent No. 5834290
: GENERAL INFORMATION:
: APPLICANT: Egelrud, Torbjorn
: APPLICANT: Hansson, Lennart
: TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
: TITLE OF INVENTION: Enzyme (SCCE)
: NUMBER OF SEQUENCES: 17
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: White & Case, Patent Department
: STREET: 1155 Avenue of the Americas
: CITY: New York
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 10036-2787
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-557-146-2

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
 | | | | |
Db 4 SLLPLQL 12

RESULT 3
US-08-824-874-3
; Sequence 3, Application US/08824874
; Patent No. 5962300
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE: Filed Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

; LIBRARY: GenBank
; CLONE: 532504
; US-08-824-874-3
Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 SLLPLQL 9
 | | | | |
Db 4 SLLPLQL 12

RESULT 4
US-09-154-344-2
; Sequence 2, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-154-344-2

Query Match 100.0%; Score 40; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
 | | | | |
Db 4 SLLPLQL 12

RESULT 5
US-08-930-188-2
; Sequence 2, Application US/08930188
; Patent No. 6093397
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.

```
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,188
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-930-188-2

Query Match 100.0%; Score 40; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
Db 4 SLLPLQL 12

RESULT 6
US-09-210-084-3
; Sequence 3, Application US/09210084
; Patent No. 6197511
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,084
; FILING DATE:
```

```
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 532504
; US-09-210-084-3

Query Match 100.0%; Score 40; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
Db 4 SLLPLQL 12

RESULT 7
PCT-US96-04294-2
; Sequence 2, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

; MOLECULE TYPE: protein
PCT-US96-04294-2

Query Match 100.0%; Score 40; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 1;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
| | | | |
Db 4 SLLPLQL 12

RESULT 8

US-09-502-600-33
; Sequence 33, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 33
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

; OTHER INFORMATION: Residues 5-13 of the SCCE protein
US-09-502-600-33
Query Match 90.0%; Score 36; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
| | | | |
Db 1 LLLPLQL 8

RESULT 9

US-09-502-600-35
; Sequence 35, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 35
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

; OTHER INFORMATION: Residues 6-14 of the SCCE protein
US-09-502-600-35
Query Match 80.0%; Score 32; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LLPLQL 9
| | | | |
Db 1 LLPLQL 7

RESULT 10

US-09-502-600-116
; Sequence 116, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 116
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

; OTHER INFORMATION: Residues 2-10 of the SCCE protein
US-09-502-600-116
Query Match 80.0%; Score 32; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLPLQ 7
| | | | |
Db 3 SLLPLQ 9

RESULT 11

US-09-177-249-13
; Sequence 13, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:

; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramin
; APPLICANT: Margossian, Linda
; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
; FILE REFERENCE: 023070-08612005
; CURRENT FILING DATE: 1998-10-22
; CURRENT APPLICATION NUMBER: US/09/177,249
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-177-249-13
Query Match 77.5%; Score 31; DB 4; Length 307;
Best Local Similarity 77.8%; Pred. No. 72;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
| | | | |
Db 60 SLLPLSFL 68

RESULT 12

US-08-383-621-5
; Sequence 5, Application US/08383621
; Patent No. 5951972

GENERAL INFORMATION:
APPLICANT: Daley, Michael J.
APPLICANT: Buckwalter, Brian L.
APPLICANT: Cady, Susan M.
APPLICANT: Shieh, Hong-Wing
APPLICANT: Bohlen, Peter
APPLICANT: Seddon, Andrew P.
TITLE OF INVENTION: Stabilization Of Somatotropins And Other
TITLE OF INVENTION: Proteins By Modification Of Cysteine Residues
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dr. Estelle J. Tsevdos
STREET: 1937 West Main Street, P.O. Box 60
CITY: Stamford
STATE: Connecticut
COUNTRY: U.S.A.
ZIP: 06904-0060
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/383,621
FILING DATE: 06-FEB-1995
CLASSIFICATION: 514
PRIOR APPLICATION NUMBER: US 07/766,142
FILING DATE: 23-SEP-1991
ATTORNEY/AGENT INFORMATION:
NAME: Tsevdos, Estelle J.
REGISTRATION NUMBER: 31,145
REFERENCE/DOCKET NUMBER: 31,278-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 203-321-2736
TELEFAX: 203-321-2971
TELEX: 203-710-474-4059
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 135 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-383-621-5

Query Match 75.0%; Score 30; DB 2; Length 135;
Best Local Similarity 77.8%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 16 SLLDLQLL 24

RESULT 13
US-08-459-906-5
Sequence 5, Application US/08459906
Patent No. 6010999
GENERAL INFORMATION:
APPLICANT: Daley, Michael J.
APPLICANT: Buckwalter, Brian L.
APPLICANT: Cady, Susan M.
APPLICANT: Shieh, Hong-Wing
APPLICANT: Bohlen, Peter
APPLICANT: Seddon, Andrew P.
TITLE OF INVENTION: Stabilization Of Somatotropins And Other
TITLE OF INVENTION: Proteins by Modification Of Cysteine Residues
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: American Cyanamid Company
STREET: One Cyanamid Plaza
CITY: Wayne
STATE: New Jersey

COUNTRY: U.S.A.
ZIP: 07470-8426
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/459,906
FILING DATE: 02-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Webster, Darryl L.
REGISTRATION NUMBER: 34,276
REFERENCE/DOCKET NUMBER: 31,278-03
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201-831-3247
TELEFAX: 201-831-3305
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 135 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-459-906-5

Query Match 75.0%; Score 30; DB 3; Length 135;
Best Local Similarity 77.8%; Pred. No. 47;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 16 SLLDLQLL 24

RESULT 14
US-08-677-049-6
Sequence 6, Application US/08677049
Patent No. 5858707
GENERAL INFORMATION:
APPLICANT: Guimaraes, M. Jorge
APPLICANT: Bazan, J. Fernando
APPLICANT: McLanahan, Terrill K.
APPLICANT: Zlotnik, Albert
TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
TITLE OF INVENTION: NUCLEIC ACIDS; ANTIBODIES
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/677,049
FILING DATE: 03-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/000,788
FILING DATE: 03-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: DX0511
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-852-9196
TELEFAX: 415-496-1200

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
NAME/KEY: Region
LOCATION: 151..180
OTHER INFORMATION: /note= "Encompasses TM 4 of Figure 4"
OTHER INFORMATION: 4*
FEATURE:
NAME/KEY: Region
LOCATION: 366..428
OTHER INFORMATION: /note= "Encompasses TM 9 and TM 10"
OTHER INFORMATION: of Figure 4"

US-08-677-049-6

Query Match 75.0%; Score 30; DB 2; Length 463;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
||||| :
Db 269 SLLPLMLV 277

RESULT 15

US-08-676-279-57

Sequence 57, Application US/08676279

Patent No. 5869247

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: MACROPHAGE NUCLEOTIDE SEQUENCE

NUMBER OF SEQUENCES: 63

COMPUTER READABLE FORM:

MEDIUM TYPE: IBM PC compatible

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/676,279

FILING DATE:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/GB95/00095

APPLICATION NUMBER: GB 9400929.7

FILING DATE: 19-JAN-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9422021.7

FILING DATE: 31-OCT-1994

INFORMATION FOR SEQ ID NO: 57:

SEQUENCE CHARACTERISTICS:

LENGTH: 549 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-676-279-57

Query Match 75.0%; Score 30; DB 2; Length 549;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
||||| :
Db 435 SLLPLMLV 443

RESULT 16

US-07-777-715-7

Sequence 7, Application US/07777715

Patent No. 5273889

GENERAL INFORMATION:

APPLICANT: Potter, Andrew
APPLICANT: Campos, Manuel
APPLICANT: Hughes, Huw P.A.
TITLE OF INVENTION: CYTOKINE-LEUKOTOXIN GENE FUSIONS AND
TITLE OF INVENTION: USES THEREOF
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Morrison & Foerster
STREET: 545 Middlefield Road, Suite 200
CITY: Menlo Park
STATE: California
COUNTRY: USA
ZIP: 94025
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/777,715
FILING DATE: 19911016
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Robins, Roberta L.
REGISTRATION NUMBER: 33,208
REFERENCE/DOCKET NUMBER: 29310-2001320
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-327-7250
TELEFAX: 415-327-2951
TELEX: 706141
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1098 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-777-715-7

Query Match 75.0%; Score 30; DB 1; Length 1098;
Best Local Similarity 77.8%; Pred. No. 4.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
||||| :
Db 23 SLLPLQLL 31

RESULT 17

US-08-170-126-2

Sequence 2, Application US/08170126

Patent No. 5594107

GENERAL INFORMATION:

APPLICANT: POTTER, ANDREW

APPLICANT: CAMPOS, MANUEL

APPLICANT: HUGHES, HUW P.A.

TITLE OF INVENTION: CYTOKINE-CYTOTOXIN GENE FUSIONS AND USES

TITLE OF INVENTION: THEREOF

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: REED & ROBINS

STREET: 635 BRYANT STREET

CITY: PALO ALTO

STATE: CALIFORNIA

COUNTRY: UNITED STATES OF AMERICA

ZIP: 94301

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/170,126

FILING DATE: 20-DEC-1993


```

; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/777,715
; PRIOR APPLICATION DATA: US 07/571,301
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0013.21
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-170-126-2

Query Match 75.0%; Score 30; DB 1; Length 1098;
Best Local Similarity 77.8%; Pred. No. 4.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 23 SLLDLQL 31

RESULT 18
US-08-954-418-2
; Sequence 2, Application US/08954418
; Patent No. 6096320
; GENERAL INFORMATION:
; APPLICANT: POTTER, ANDREW
; APPLICANT: CAMPOS, MANUEL
; APPLICANT: HUGHES, HOW P.A.
; TITLE OF INVENTION: CYTOKINE-CYTOTOXIN GENE FUSIONS AND USES
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: REED & ROBINS
; STREET: 635 BRYANT STREET
; CITY: PALO ALTO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94301
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/954,418
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/170,126
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/571,301
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 9000-0013.21
; TELEPHONE: (415) 617-8999
; TELEFAX: (415) 327-3231
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1098 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

```

```

US-08-954-418-2

Query Match 75.0%; Score 30; DB 3; Length 1098;
Best Local Similarity 77.8%; Pred. No. 4.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 23 SLLDLQL 31

RESULT 19
US-08-753-007A-24
; Sequence 24, Application US/08753007A
; Patent No. 6074841
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.
; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/753,007A
; FILING DATE: 19-NOV-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/699,591
; FILING DATE: 19-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Fasse, J. Peter
; REGISTRATION NUMBER: 32,983
; REFERENCE/DOCKET NUMBER: 07334/022001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX:
; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 21 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-753-007A-24

Query Match 72.5%; Score 29; DB 3; Length 21;
Best Local Similarity 66.7%; Pred. No. 10;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 7 ALLPLALL 15

RESULT 20
US-09-398-496-24
; Sequence 24, Application US/09398496
; Patent No. 6133423
; GENERAL INFORMATION:
; APPLICANT: Gearing, David P.
; APPLICANT: Busfield, Samantha J.

```

;; TITLE OF INVENTION: DON-1 GENE AND POLYPEPTIDES
;; TITLE OF INVENTION: AND USES THEREFOR
;; NUMBER OF SEQUENCES: 33
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Fish & Richardson P.C.
;; STREET: 225 Franklin Street
;; CITY: Boston
;; STATE: MA
;; COUNTRY: US
;; ZIP: 02110-2804
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Diskette
;; COMPUTER: IBM Compatible
;; OPERATING SYSTEM: DOS
;; SOFTWARE: FastSeq Version 2.0
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/398,496
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/753,007
;; FILING DATE: 19-NOV-1996
;; APPLICATION NUMBER: 08/699,591
;; FILING DATE: 19-AUG-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Fasse, J. Peter
;; REGISTRATION NUMBER: 32,983
;; REFERENCE/DOCKET NUMBER: 07334/022001
;; TELEPHONE: 617-542-5070
;; TELEFAX: 617-542-8906
;; TELEX:
;; INFORMATION FOR SEQ ID NO: 24:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 21 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: not relevant
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; US-09-398-496-24

Query Match 72.5%; Score 29; DB 4; Length 21;
Best Local Similarity 66.7%; Pred. No. 10;
Matches 6; Conservative 2; Mismatches 1; Indels 1; Gaps 0;

Qy 1 SLLPLQL 9
Db 7 ALLPLALL 15

RESULT 21
US-08-977-378-22
; Sequence 22, Application US/08977378
; Patent No. 6309842
; GENERAL INFORMATION:
; APPLICANT: Dower, William J.
; Gates, Christopher M.
; Heinkel, Gregory L.
; Lalonde, Guy
; Mattheakis, Larry C.
; Paddon, Christopher J.
; Schatz, Peter J.
; TITLE OF INVENTION: Use of Modified Tethers in Screening
; Compound Libraries
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834
; COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC Compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/977,378
;; FILING DATE: 24-No. 6309842-1997
;; CLASSIFICATION: <Unknown>
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/758,307
;; FILING DATE: 03-DEC-1996
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Liebeschuetz, Joe
;; REGISTRATION NUMBER: 37,505
;; REFERENCE/DOCKET NUMBER: 16528A-018010US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; INFORMATION FOR SEQ ID NO: 22:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 22 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: <Unknown>
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
;; SEQUENCE DESCRIPTION: SEQ ID NO: 22:
US-08-977-378-22

Query Match 72.5%; Score 29; DB 4; Length 22;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
Db 7 ALLPLALL 15

RESULT 22
US-08-630-172-40
; Sequence 40, Application US/08630172
; Patent No. 6060054
; GENERAL INFORMATION:
; APPLICANT: Staerz, Uwe
; TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
; TITLE OF INVENTION: LYMPHOCYTE VETO
; NUMBER OF SEQUENCES: 41
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross & McIntosh
; STREET: 1700 Lincoln Street, 35th Floor
; CITY: Denver
; STATE: Colorado
; COUNTRY: U.S.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/630,172
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Connell, Gary J.
; REGISTRATION NUMBER: 32,020
; REFERENCE/DOCKET NUMBER: 2879-36
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (303) 863-9700
; TELEFAX: (303) 863-0223
; INFORMATION FOR SEQ ID NO: 40:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 26 amino acids
; TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..26
OTHER INFORMATION: /label= leader
US-08-630-172-40

Query Match 72.5%; Score 29; DB 3; Length 26;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
Db 15 LLLPLSL 22

RESULT 23
US-09-375-419-40
Sequence 40, Application US/09375419
Patent No. 6264950

GENERAL INFORMATION:
APPLICANT: Staerz, Uwe
TITLE OF INVENTION: NOVEL PRODUCT AND PROCESS FOR T
TITLE OF INVENTION: LYMPHOCYTE VETO
NUMBER OF SEQUENCES: 41
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, 35th Floor
CITY: Denver
STATE: Colorado
COUNTRY: U.S.
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/375,419
FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/630,172
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2879-36
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223

INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 26 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..26
OTHER INFORMATION: /label= leader
US-09-375-419-40

Query Match 72.5%; Score 29; DB 4; Length 26;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
Db 15 LLLPLSL 22

RESULT 24
US-08-339-152A-19
Sequence 19, Application US/08339152A
Patent No. 5643726
GENERAL INFORMATION:
APPLICANT: Tanzi, Rudolph E.
APPLICANT: Kovacs, Dora M.
TITLE OF INVENTION: Methods For Modulating Transcription
TITLE OF INVENTION: From The Amyloid -Protein Precursor (APP) Promoter
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Ave., NW, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/339,152A
FILING DATE: 10-NOV-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 0609.4120000
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
TELEX:

INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 190 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-339-152A-19

Query Match 72.5%; Score 29; DB 1; Length 190;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
Db 1 LLLPLSL 8

RESULT 25
US-08-007-999B-6
Sequence 6, Application US/08007999B
Patent No. 5851787
GENERAL INFORMATION:
APPLICANT: Wasco, Wilma
APPLICANT: Bupp, Keith
APPLICANT: Magendantz, Margaret
APPLICANT: Tanzi, Rudolph
APPLICANT: Solomon, Frank
TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
STREET: 1100 New York Ave., NW
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/007,999B
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, G. Kevin
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-007-999B-6

Query Match 72.5%; Score 29; DB 2; Length 190;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
 |||||:1
Db 1 LLLPLSL 8

RESULT 26

US-08-689-276A-6
; Sequence 6, Application US/08689276A
; Patent No. 5891991
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,276A
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: JORGE A. GOLDSTEIN

; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.3520003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 190 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-689-276A-6

Query Match 72.5%; Score 29; DB 2; Length 190;
Best Local Similarity 75.0%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
 |||||:1
Db 1 LLLPLSL 8

RESULT 27

US-07-940-605A-12
; Sequence 12, Application US/07940605A
; Patent No. 5540926
; GENERAL INFORMATION:
; APPLICANT: ARUFFO, ALEJANDRO
; APPLICANT: HOLLENBAUGH, DIANE
; APPLICANT: LEDBETTER, JEFFREY A.
; TITLE OF INVENTION: SOLUBLE LIGANDS FOR CD40
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/940,605A
; FILING DATE: 04-SEP-1992
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 5624-184
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-07-940-605A-12

Query Match 72.5%; Score 29; DB 1; Length 235;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
 |||||:1
Db 7 ALLPLAL 15

RESULT 28
US-08-690-096-12
; Sequence 12, Application US/08690096
; Patent No. 5945513
; GENERAL INFORMATION:
; APPLICANT: ARUFFO, ALEJANDRO
; APPLICANT: HOLLENBAUGH, DIANE
; APPLICANT: LEBETTER, JEFFREY A.
; TITLE OF INVENTION: SOLUBLE LIGANDS FOR CD40
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/690,096
; FILING DATE: 31-JUL-1996
; CLASSIFICATION: 536
; PRIOR APPLICATION NUMBER: US 07/940,605
; FILING DATE: 04-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Misrock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 5624-184
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212 790-9090
; TELEFAX: 212 869-8864/9741
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 235 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-690-096-12

Query Match 72.5%; Score 29; DB 2; Length 235;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 7 ALLPLALL 15

RESULT 29
US-08-765-907A-8
; Sequence 8, Application US/08765907A.
; Patent No. 6352839
; GENERAL INFORMATION:
; APPLICANT: BLANC, Veronique
; APPLICANT: THIBAUT, Denis
; APPLICANT: BAMA-JACQUES, Nathalie
; APPLICANT: BLANCHE, Francis
; APPLICANT: COUZET, Joel
; APPLICANT: BARRIERE, Jean-Claude
; APPLICANT: DEBUSSCHE, Laurent
; APPLICANT: FAMECHON, Alain
; APPLICANT: PARIS, Jean-Marc
; APPLICANT: DUTRUC-ROSSET, Gilles
; TITLE OF INVENTION: Streptogramins And Method For Preparing Same By
; TITLE OF INVENTION: Mutasythesis
; FILE REFERENCE: Streptogramin genes
; CURRENT APPLICATION NUMBER: US/08/765,907A

; CURRENT FILING DATE: 1997-03-20
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Streptomyces pristinaespiralis
US-08-765-907A-8

Query Match 72.5%; Score 29; DB 4; Length 356;
Best Local Similarity 85.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 7
Db 158 SLVPLQL 164

RESULT 30
US-08-751-512-8
; Sequence 8, Application US/08751512
; Patent No. 6001962
; GENERAL INFORMATION:
; APPLICANT: Ramer, J. Kevin
; APPLICANT: Williams, Lewis T.
; TITLE OF INVENTION: Modified FAS Ligands
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
; STREET: Two Embarcadero Center, 8th Floor
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/751,512
; FILING DATE: 15-NOV-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy, Matthew B.
; REGISTRATION NUMBER: 39,787
; REFERENCE/DOCKET NUMBER: 02307K-07100US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-576-0200
; TELEFAX: 415-576-0300
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 376 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-751-512-8

Query Match 72.5%; Score 29; DB 3; Length 376;
Best Local Similarity 66.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 7 ALLPLALL 15

RESULT 31
US-09-554-225-1
; Sequence 1, Application US/09554225
; Patent No. 6252056
; GENERAL INFORMATION:
; APPLICANT: FUKUSHIMA, DAICHI

```

; APPLICANT: NAKADE, SHINJI
; APPLICANT: HAGA, HISANORI
; TITLE OF INVENTION: HUMAN LYSOPHOSPHATIDIC ACID RECEPTOR AND USE THEREOF
; FILE REFERENCE: Q59116
; CURRENT APPLICATION NUMBER: US/09/554.225
; CURRENT FILING DATE: 2000-05-11
; PRIOR APPLICATION NUMBER: PCT/JP98/05047
; PRIOR FILING DATE: 1998-11-10
; PRIOR APPLICATION NUMBER: JP P. Hei 9-307749
; PRIOR FILING DATE: 1997-11-11
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 1
; LENGTH: 419
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-554-225-1

Query Match 72.5%; Score 29; DB 4; Length 419;
Best Local Similarity 87.5%; Pred. No. 2.5e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLPLQI 8
   |||||
Db 75 SLNPLQI 82

RESULT 32
US-08-637-823B-4
; Sequence 4, Application US/08637823B
; Patent No. 6184031
; GENERAL INFORMATION:
; APPLICANT: Gros, Philippe
; APPLICANT: Skamene, Emil
; TITLE OF INVENTION: DNA SEQUENCES THAT ENCODE A NATURAL
; TITLE OF INVENTION: RESISTANCE TO INFECTION WITH INTRACELLULAR PARASITES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KLAUBER & JACKSON
; STREET: 411 Hackensack Ave
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,823B
; FILING DATE: 05/08/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson, David A.
; REGISTRATION NUMBER: 26,742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487 5800
; TELEFAX: 201 343 1684
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-637-823B-4

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,823B
; FILING DATE: 05/08/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson, David A.
; REGISTRATION NUMBER: 26,742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487 5800
; TELEFAX: 201 343 1684
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-637-823B-4

Query Match 72.5%; Score 29; DB 4; Length 483;
Best Local Similarity 66.7%; Pred. No. 2.9e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SLLPLQI 9
   ||||| :|
Db 75 SLNPLQI 82

RESULT 34
US-08-903-139B-7
; Sequence 7, Application US/08903139B
; Patent No. 6114118
; GENERAL INFORMATION:
; APPLICANT: Joe W. Templeton, Jianwei Peng, L. Garry Adams,
; APPLICANT: Erwin Schurr, Philippe Gros, Donald S. Davis and Roger Smith
; TITLE OF INVENTION: METHOD OF IDENTIFICATION OF ANIMALS
; TITLE OF INVENTION: RESISTANT OR SUSCEPTIBLE TO DISEASES SUCH AS RUMINANT
; TITLE OF INVENTION: BRUCellosis, TUBERCULOSIS, PARATUBERCULOSIS AND SALMONELLOSIS
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
; STREET: 1177 West Loop South, 10th Floor
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77027-9095
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/903,139B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031,443
FILING DATE: September 20, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Krieger, Paul E.
REGISTRATION NUMBER: 25,886
REFERENCE/DOCKET NUMBER: 00162-3/V96171US
TELEPHONE: 713-850-0909
TELEFAX: 713-850-0165
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 547 amino acids
TYPE: amino acids
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-903-139B-7

Query Match 72.5%; Score 29; DB 3; Length 547;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
Db 433 SLLPFAVL 441

RESULT 35
US-08-676-279-50
Sequence 50, Application US/08676279
Patent No. 5869247
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: MACROPHAGE NUCLEOTIDE SEQUENCE
NUMBER OF SEQUENCES: 63
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,279
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/00095
APPLICATION NUMBER: GB 9400929.7
FILING DATE: 19-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9422021.7
FILING DATE: 31-OCT-1994
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-676-279-50

Query Match 72.5%; Score 29; DB 2; Length 548;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
Db 433 SLLPFAVL 441

RESULT 36
US-08-903-139B-8
Sequence 8, Application US/08903139B
Patent No. 611418
GENERAL INFORMATION:
APPLICANT: Joe W. Templeton, Jianwei Feng, L. Garry Adams,
APPLICANT: Erwin Schurr, Philippe Gros, Donald S. Davis and Roger Smith
TITLE OF INVENTION: METHOD OF IDENTIFICATION OF ANIMALS
RESISTANT OR SUSCEPTIBLE TO DISEASES SUCH AS RUMINANT
TITLE OF INVENTION: BRUCELLOSIS, TUBERCULOSIS, PARATUBERCULOSIS AND SALMONELLA
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
STREET: 1177 West Loop South, 10th Floor
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77027-9095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/903,139B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031,443
FILING DATE: September 20, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Krieger, Paul E.
REGISTRATION NUMBER: 25,886
REFERENCE/DOCKET NUMBER: 00162-3/V96171US
TELEPHONE: 713-850-0909
TELEFAX: 713-850-0165
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-903-139B-8

Query Match 72.5%; Score 29; DB 3; Length 548;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
Db 433 SLLPFAVL 441

RESULT 37
US-08-903-139B-9
Sequence 9, Application US/08903139B
Patent No. 611418
GENERAL INFORMATION:
APPLICANT: Joe W. Templeton, Jianwei Feng, L. Garry Adams,
APPLICANT: Erwin Schurr, Philippe Gros, Donald S. Davis and Roger Smith
TITLE OF INVENTION: METHOD OF IDENTIFICATION OF ANIMALS
RESISTANT OR SUSCEPTIBLE TO DISEASES SUCH AS RUMINANT
TITLE OF INVENTION: BRUCELLOSIS, TUBERCULOSIS, PARATUBERCULOSIS AND SALMONELLA
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
STREET: 1177 West Loop South, 10th Floor
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77027-9095

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/903,139B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031,443
FILING DATE: September 20, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Krieger, Paul E.
REGISTRATION NUMBER: 25,886
REFERENCE/DOCKET NUMBER: 00162-3/V96171US
TELEPHONE: 713-850-0909
TELEFAX: 713-850-0165
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-903-139B-9

Query Match 72.5%; Score 29; DB 3; Length 548;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 433 SLLPLFAVL 441

RESULT 38
US-08-903-139B-28
Sequence 28, Application US/08903139B
Patent No. 6114118
GENERAL INFORMATION:
APPLICANT: Joe W. Templeton, Jianwei Peng, L. Garry Adams,
APPLICANT: Erwin Schurr, Philippe Gros, Donald S. Davis and Roger Smith
TITLE OF INVENTION: METHOD OF IDENTIFICATION OF ANIMALS
TITLE OF INVENTION: RESISTANT OR SUSCEPTIBLE TO DISEASES SUCH AS RUMINANT
TITLE OF INVENTION: BRUCellosis, TUBERCULOSIS, PARATUBERCULOSIS AND SALMONELLOSIS
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pravel, Hewitt, Kimball & Krieger
STREET: 1177 West Loop South, 10th Floor
CITY: Houston
STATE: TX
COUNTRY: USA
ZIP: 77027-9095
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/903,139B
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/031,443
FILING DATE: September 20, 1996
ATTORNEY/AGENT INFORMATION:
NAME: Krieger, Paul E.
REGISTRATION NUMBER: 25,886
REFERENCE/DOCKET NUMBER: 00162-3/V96171US
TELEPHONE: 713-850-0909
TELEFAX: 713-850-0165

INFORMATION FOR SEQ ID NO: 28:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: peptide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-903-139B-28

Query Match 72.5%; Score 29; DB 3; Length 548;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 433 SLLPLFAVL 441

RESULT 39
US-08-637-823B-25
Sequence 25, Application US/08637823B
Patent No. 6184031
GENERAL INFORMATION:
APPLICANT: Gros, Philippe
APPLICANT: Skamene, Emil
TITLE OF INVENTION: DNA SEQUENCES THAT ENCODE A NATURAL
TITLE OF INVENTION: RESISTANCE TO INFECTION WITH INTRACELLULAR PARASITES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: KLAUBER & JACKSON
STREET: 411 Hackensack Ave
CITY: Hackensack
STATE: New Jersey
COUNTRY: U.S.A.
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,823B
FILING DATE: 05/08/96
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson, David A.
REGISTRATION NUMBER: 26,742
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487 5800
TELEFAX: 201 343 1684
TELEX:
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 548 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-637-823B-25

Query Match 72.5%; Score 29; DB 4; Length 548;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
Db 433 SLLPLFAVL 441

RESULT 40
US-08-637-823B-30
Sequence 30, Application US/08637823B
Patent No. 6184031
GENERAL INFORMATION:


```
; APPLICANT: Gros, Philippe
; TITLE OF INVENTION: DNA SEQUENCES THAT ENCODE A NATURAL
; TELECOMMUNICATION INFORMATION:
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: KLAUBER & JACKSON
; STREET: 411 Hackensack Ave
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/637,823B
; FILING DATE: 05/08/96
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson, David A.
; REGISTRATION NUMBER: 26,742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487 5800
; TELEFAX: 201 343 1684
; TELEX:
; INFORMATION FOR SEQ ID NO: 30:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 550 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-637-823B-30

Query Match 72.5%; Score 29; DB 4; Length 550;
Best Local Similarity 66.7%; Pred. No. 3.3e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 SLLPLQIL 9
Db 436 SLLPLFAVL 444

RESULT 41
US-08-339-152A-17
; Sequence 17, Application US/08339152A
; Patent No. 5643726
; GENERAL INFORMATION:
; APPLICANT: Tanzil, Rudolph E.
; REGISTRATION NUMBER: 36,688
; TITLE OF INVENTION: Methods For Modulating Transcription
; TELECOMMUNICATION INFORMATION:
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; TITLE OF INVENTION: Methods For Modulating Transcription
; TELECOMMUNICATION INFORMATION:
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
```

```
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 0609.4120000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; TELEX:
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 634 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-339-152A-17

Query Match 72.5%; Score 29; DB 1; Length 634;
Best Local Similarity 75.0%; Pred. No. 3.9e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQIL 9
Db 1 LLLPLSLL 8

RESULT 42
US-08-339-152A-16
; Sequence 16, Application US/08339152A
; Patent No. 5643726
; GENERAL INFORMATION:
; APPLICANT: Tanzil, Rudolph E.
; REGISTRATION NUMBER: 36,688
; TITLE OF INVENTION: Methods For Modulating Transcription
; TELECOMMUNICATION INFORMATION:
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; TITLE OF INVENTION: Methods For Modulating Transcription
; TELECOMMUNICATION INFORMATION:
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; TITLE OF INVENTION: Methods For Modulating Transcription
; TELECOMMUNICATION INFORMATION:
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; TITLE OF INVENTION: Methods For Modulating Transcription
; TELECOMMUNICATION INFORMATION:
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Ave., NW, Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/339,152A
; FILING DATE: 10-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
```

```
Query Match 72.5%; Score 29; DB 1; Length 653;
Best Local Similarity 75.0%; Pred. No. 4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 LLLPLQIL 9
Db 21 LLLPLSLL 28
```

RESULT 43

US-08-007-999B-3
; Sequence 3, Application US/08007999B
; Patent No. 5851787
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox, P.L.L.C.
; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/007,999B
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Townsend, G. Kevin
; REGISTRATION NUMBER: 34,033
; REFERENCE/DOCKET NUMBER: 0609.3520002/JAG/GKT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 653 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-007-999B-3

Query Match 72.5% Score 29; DB 2; Length 653;
Best Local Similarity 75.0% Pred. No. 4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
Db 21 LLLPLSL 28

RESULT 44

US-08-689-276A-3
; Sequence 3, Application US/08689276A
; Patent No. 5891991
; GENERAL INFORMATION:
; APPLICANT: Wasco, Wilma
; APPLICANT: Bupp, Keith
; APPLICANT: Magendantz, Margaret
; APPLICANT: Tanzi, Rudolph
; APPLICANT: Solomon, Frank
; TITLE OF INVENTION: AMYLOID PRECURSOR-LIKE PROTEIN AND USES
; TITLE OF INVENTION: THEREOF
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.

; STREET: 1100 New York Ave., NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/689,276A
; FILING DATE: 06-AUG-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,999
; FILING DATE: 21-JAN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,642
; FILING DATE: 20-APR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/930,022
; FILING DATE: 17-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: JORGE A. GOLDSTEIN
; REGISTRATION NUMBER: 29,021
; REFERENCE/DOCKET NUMBER: 0609.3520003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)371-2571
; TELEFAX: (202)371-2540
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 653 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-689-276A-3

Query Match 72.5% Score 29; DB 2; Length 653;
Best Local Similarity 75.0% Pred. No. 4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
Db 21 LLLPLSL 28

RESULT 45

US-08-658-136-5
; Sequence 5, Application US/08658136
; Patent No. 6071717
; GENERAL INFORMATION:
; APPLICANT: KLINGER, KATHERINE W
; APPLICANT: LANDES, GREGORY M
; APPLICANT: BURN, TIMOTHY C
; APPLICANT: CONNORS, TIMOTHY D
; APPLICANT: DACKOWSKI, WILLIAM
; APPLICANT: GERMINO, GREGORY
; APPLICANT: QIAN, PENG
; TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE GENE
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENZYME CORPORATION
; STREET: ONE MOUNTAIN ROAD
; CITY: FRAMINGHAM
; STATE: MASSACHUSETTS
; COUNTRY: USA
; ZIP: 01701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

; APPLICATION NUMBER: US/08/658,136
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: LASSEN, ELIZABETH
; REGISTRATION NUMBER: 31,845
; REFERENCE/DOCKET NUMBER: GEN4-17.8
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 508-872-8400
; TELEFAX: 508-872-5415
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4302 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-658-136-5

```

```

Query Match 72.5%; Score 29; DB 3; Length 4302;
Best Local Similarity 62.5%; Pred. No. 2.9e+03;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy 2 LLLPLQIL 9
Db 2153 VVLPQLVL 2160

```

```

Search completed: November 6, 2002, 12:06:29
Job time : 10.1111 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 23.7778 seconds
(without alignments)
42.042 Million cell updates/sec

Title: US-09-905-083-36
Perfect score: 40
Sequence: 1 SLLPLQLIL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	40	100.0	9	AAE08241	Human stratum corn
2	40	100.0	136	ABG23378	Novel human diago
3	40	100.0	253	AAR67888	Human stratum corn
4	40	100.0	253	AAW05383	Human amyloid prec
5	40	100.0	257	AAE21326	Human HSCCE. Homo
6	36	90.0	9	AAE08238	Human stratum corn
7	34	85.0	61	AAQ12472	Human polypeptide
8	32	80.0	9	AAE08240	Human stratum corn
9	32	80.0	9	AAE08320	Human stratum corn
10	32	80.0	71	AAM89706	Human immune/haema
11	32	80.0	73	ABG19347	Novel human diago

12	32	80.0	267	18	AAW22303	Rat CRTI. Rattus
13	31	77.5	46	21	AAB38062	Fragment of human
14	31	77.5	56	22	AAW79694	Human protein SEQ
15	31	77.5	89	22	ABG12769	Novel human diago
16	31	77.5	89	22	ABG13602	Novel human diago
17	31	77.5	142	22	AAE63580	Human gastric canc
18	31	77.5	156	22	AAE63578	Human gastric canc
19	31	77.5	159	22	AAE63582	Human gastric canc
20	31	77.5	189	22	ABE71026	Drosophila melanog
21	31	77.5	211	21	AAB38058	Fragment of human
22	31	77.5	246	21	AAB24037	Human PRO1555 prot
23	31	77.5	246	21	AAE99438	Human PRO1555 (UNQ
24	31	77.5	246	22	AAE66187	protein of the inv
25	31	77.5	278	22	AAU08753	Human insulin-like
26	31	77.5	316	22	ABE62737	Drosophila melanog
27	31	77.5	330	22	ABG01683	Novel human diago
28	31	77.5	390	22	AAU08755	Human insulin-like
29	31	77.5	393	22	AAU17125	Novel signal trans
30	31	77.5	585	20	AAE23643	protein encoded by
31	31	77.5	770	22	ABG25713	Novel human diago
32	31	77.5	925	22	ACG92740	C glutamicum prote
33	31	77.5	1062	22	ABG01394	Novel human diago
34	30	75.0	23	22	ABE43858	Peptide #11364 enc
35	30	75.0	23	22	AAE64837	Human brain expres
36	30	75.0	23	22	AAE37771	Peptide #11808 enc
37	30	75.0	37	22	AAU01749	Human secreted pro
38	30	75.0	40	22	ABE38810	Peptide #6316 enco
39	30	75.0	40	22	ABE23836	Protein #5835 enco
40	30	75.0	40	22	AAE59450	Human brain expres
41	30	75.0	40	22	AAE72012	Human bone marrow
42	30	75.0	40	22	AAE19406	Peptide #5840 enco
43	30	75.0	40	22	AAE32284	Peptide #6321 enco
44	30	75.0	70	21	AAE18531	Zea mays protein f
45	30	75.0	83	22	AAE84905	Human immune/haema

ALIGNMENTS

RESULT 1

AAE08241
ID AAE08241 standard; peptide; 9 AA.

AC AAE08241;

XX
DT 01-NOV-2001 (first entry)

XX Human stratum corneum chymotrypsin enzyme peptide #6 (residues 4-12).

DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antitense therapy; malignant hyperplasia.

XX Homo sapiens.

OS
XX WO200159158-A1.

XX 16-AUG-2001.

XX 07-FEB-2001; 2001WO-US03977.

XX 11-FEB-2000; 2000US-0502600.

XX (UYAR-) UNIV ARKANSAS.

XX O'brien TJ;

XX WPI; 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -

XX Claim 25; Page 103; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.

XX Sequence 9 AA;

Query Match 100.0%; Score 40; DB 22; Length 9;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

Db 1 SLLPLQL 9

RESULT 2

ABG23378
ID ABG23378 standard; Protein; 136 AA.

XX AC ABG23378;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #23369.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
XX food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI; 2001-639362/73.

XX DR N-PSDB; AAS87565.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

XX PS Claim 20; SEQ ID No 53737; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 136 AA;

Query Match 100.0%; Score 40; DB 22; Length 136;

Best Local Similarity 100.0%; Pred. No. 2.7;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

Db 10 SLLPLQL 18

RESULT 3

AAR6788
ID AAR6788 standard; Protein; 253 AA.

XX AC AAR6788;

XX DT 09-AUG-1995 (first entry)

XX DE Human stratum corneum chymotrophic recombinant enzyme (SCCE).

XX KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
XX callosities; keratosis pilaris; ichthyoses; eczema.

XX OS Homo sapiens.

XX PN WO9500651-A.

XX PD 05-JAN-1995.

XX PF 20-JUN-1994; 94WO-1B00166.

XX PR 18-JUN-1993; 93DK-0000725.

XX PA (SYMB-) SYMBICOM AB.

XX PI Egelrud T, Hansson L;

XX DR WPI; 1995-052088/07.

XX DR N-PSDB; AAQ81203.

XX PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme
PT - and related vectors, transformed cells and polypeptides, and
PT useful for treating skin disorders, e.g. acne or psoriasis, and
PT for identification of specific inhibitors.

XX PS Disclosure; Page 97; 137pp; English.

XX CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
CC and skin care products, especially to treat and prevent acne,
CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
CC produced recombinantly following mammal, insect, plant, or
CC microorganism transformation with plasmid pS507.

XX Sequence 253 AA;

Query Match 100.0%; Score 40; DB 16; Length 253;

Best Local Similarity 100.0%; Pred. No. 5.2;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

Db 4 SLLPLQL 12

```
RESULT 4
AAW05383
ID AAW05383 standard; Protein; 253 AA.
XX
AC AAW05383;
XX
DT 31-DEC-1996 (first entry)
XX
DE Human amyloid precursor protein protease.
XX
KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;
KW therapy.
OS Homo sapiens.
XX
PN W09631122-A1.
XX
PD 10-OCT-1996.
XX
PF 02-APR-1996; 96WO-US04294.
XX
PR 04-APR-1995; 95US-0416257.
XX
PA (ELIL ) LILLY & CO ELI.
XX
PI Dixon EP, Johnstone EM, Little SP;
XX
DR WPI; 1996-464694/46.
DR N-PSDB; AAT39783.
XX
PT New isolated human amyloid precursor protein protease - used to
PT develop prods. for the treatment or diagnosis of associated
PT conditions, esp. Alzheimer's disease
XX
PS Claim 1; Page 44-45; 55pp; English.
XX
CC Human amyloid precursor protein protease (AAW05383) is involved in
CC the processing or clearance of amyloid precursor protein to form
CC beta-amyloid peptide. Its amino acid sequence was deduced from
CC a cDNA clone (AAT39783) obtd. from a human lung library. Recombinant
CC protease can be produced in transformed or transfected prokaryotic
CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
CC used to develop products for the design and testing of cpds. useful
CC for treating or preventing conditions associated with beta-amyloid
CC peptide, esp. Alzheimer's disease.
XX
SQ Sequence 253 AA;
Query Match 100.0%; Score 40; DB 17; Length 253;
Best Local Similarity 100.0%; Pred. No. 5.2;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SLLPLQL 9
Db 4 SLLPLQL 12
RESULT 5
AAB21326
ID AAB21326 standard; Protein; 257 AA.
XX
AC AAB21326;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human HSCEE.
XX
KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCEE;
KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;
KW serine protease; cytostatic; cancer; prostrate cancer.
XX
PI O'brien TJ;
```

```
OS Homo sapiens.
XX
PN W0200053776-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-CA00258.
XX
PR 11-MAR-1999; 99US-0124260.
XX
PR 01-APR-1999; 99US-0127386.
XX
PR 21-JUL-1999; 99US-0144919.
XX
PA (MOUN ) MOUNT SINAI HOSPITAL.
XX
PI Yousef GM, Diamandis EP;
XX
DR WPI; 2000-587440/55.
XX
PT New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
PT protein mediated disorders, especially cancer.
XX
PS Example 4; Fig 17; 184pp; English.
XX
CC The present sequence is human stratum corneum chymotryptic enzyme
CC (HSCEE), a member of the kallikrein multi-gene family. Kallikreins and
CC kallikrein-like proteins are a subgroup of the serine protease enzyme
CC family. They catalyse the selective cleavage of specific polypeptide
CC precursors to release peptides with potent biological activity. Nucleic
CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
CC treatment, monitoring and diagnosis of cancers, especially prostate
CC cancer. They can also be used to identify a substance that can associate
CC with or mediate the biological activity of the proteins. Antibodies can
CC be used to treat conditions mediated by the kallikrein-like proteins.
XX
SQ Sequence 257 AA;
Query Match 100.0%; Score 40; DB 21; Length 257;
Best Local Similarity 100.0%; Pred. No. 5.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 SLLPLQL 9
Db 4 SLLPLQL 12
RESULT 6
AAE08238
ID AAE08238 standard; peptide; 9 AA.
XX
AC AAE08238;
XX
DT 01-NOV-2001 (first entry)
XX
DE Human stratum corneum chymotrypsin enzyme peptide #3 (residues 5-13).
XX
KW Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.
XX
OS Homo sapiens.
XX
PN W0200159158-A1.
XX
PD 16-AUG-2001.
XX
PF 07-FEB-2001; 2001WO-US03977.
XX
PR 11-FEB-2000; 2000US-0502600.
XX
PA (UYAR-) UNIV ARKANSAS.
XX
PI O'brien TJ;
```

XX WPI; 2001-514676/56.
XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -
XX
XX Claim 25; Page 102; 127pp; English.
XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.
XX
XX Sequence 9 AA;
SQ
Query Match 90.0%; Score 36; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 LLLPLQL 9
Db 1 LLLPLQL 8
RESULT 7
AAO12472
ID AAO12472 standard; Protein; 61 AA.
XX
AC AAO12472;
XX
DT 06-NOV-2001 (first entry)
XX
XX Human polypeptide SEQ ID NO 26364.
DE
DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopolesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorders; arthritis; inflammation.
XX
OS Homo sapiens.
XX
XX WO200164835-A2.
PN
XX
PD 07-SEP-2001.
XX
XX 26-FEB-2001; 2001WO-US04927.
XX
XX 28-FEB-2000; 2000US-0515126.
PR
PR 18-MAY-2000; 2000US-0577409.
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Tang YT, Liu C, Drmanac RT;
PI
XX
XX WPI; 2001-514838/56.
DR
DR N-PSDB; AAI92403.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
PT diagnosing and treating e.g. leukaemia, inflammation and immune
PT disorders -
XX
XX Claim 20; SEQ ID NO 26364; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AAI79941-AAI93841) and
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopolesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 61 AA;
SQ
Query Match 85.0%; Score 34; DB 22; Length 61;
Best Local Similarity 87.5%; Pred. No. 17;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 LLLPLQL 9
Db 4 LLLPLQL 11
RESULT 8
AAE08240
ID AAE08240 standard; peptide; 9 AA.
XX
AC AAE08240;
XX
DT 01-NOV-2001 (first entry)
XX
XX Human stratum corneum chymotrypsin enzyme peptide #5 (residues 6-14).
DE
DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.
XX
OS Homo sapiens.
XX
XX WO200159158-A1.
PN
XX
PD 16-AUG-2001.
XX
XX 07-FEB-2001; 2001WO-US03977.
PF
XX
PR 11-FEB-2000; 2000US-0502600.
XX
XX (UYAR-) UNIV ARKANSAS.
PA
XX
XX O'brien TJ;
PI
XX
XX WPI; 2001-514676/56.
DR
XX
PT Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -
XX
XX Claim 25; Page 103; 127pp; English.
XX
XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.
XX
XX Sequence 9 AA;
SQ
Query Match 80.0%; Score 32; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 LLLPLQL 9

PR	27-SEP-2000;	2000US-0235834.	XX	Rosen CA, Barash SC, Ruben SM;
PR	27-SEP-2000;	2000US-0235836.	PI	
PR	29-SEP-2000;	2000US-0236327.	XX	WPI: 2001-483426/52.
PR	29-SEP-2000;	2000US-0236367.	DR	N-PSDB; AAK62487.
PR	29-SEP-2000;	2000US-0236368.	XX	
PR	29-SEP-2000;	2000US-0236369.	XX	
PR	29-SEP-2000;	2000US-0236370.	PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PR	02-OCT-2000;	2000US-0236802.	PT	useful for preventing, diagnosing and/or treating cancers and
PR	02-OCT-2000;	2000US-0237037.	PT	metastasis.
PR	02-OCT-2000;	2000US-0237038.	XX	
PR	02-OCT-2000;	2000US-0237039.	XX	Claim 11; SEQ ID NO 17299; 3071pp + Sequence Listing; English.
PR	02-OCT-2000;	2000US-0237040.	XX	
PR	13-OCT-2000;	2000US-0239935.	CC	AAK54951 to AAK64702 encode the human Immune/hematopoietic antigen (I)
PR	13-OCT-2000;	2000US-0239937.	CC	amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
PR	20-OCT-2000;	2000US-0240960.	CC	activity, and can be used in gene therapy and vaccine production. (I)
PR	20-OCT-2000;	2000US-0241221.	CC	proteins and polynucleotides may be used in the prevention, diagnosis and
PR	20-OCT-2000;	2000US-0241221.	CC	treatment of diseases associated with inappropriate (I) expression. For
PR	20-OCT-2000;	2000US-0241785.	CC	example, they may be used to treat disorders associated with decreased
PR	20-OCT-2000;	2000US-0241786.	CC	expression by rectifying mutations or deletions in a patient's genome
PR	20-OCT-2000;	2000US-0241787.	CC	that affect the activity of (I) by expressing inactive proteins or to
PR	20-OCT-2000;	2000US-0241808.	CC	supplement the patients own production of (I). Additionally, (I)
PR	20-OCT-2000;	2000US-0241809.	CC	polynucleotides may be used to produce the secreted (I); by inserting
PR	20-OCT-2000;	2000US-0241826.	CC	the nucleic acids into a host cell and culturing the cell to express the
PR	01-NOV-2000;	2000US-0244617.	CC	protein. (I) proteins and polynucleotides may be used to prevent,
PR	08-NOV-2000;	2000US-0246474.	CC	diagnose and treat immune/hematopoietic-related diseases, especially
PR	08-NOV-2000;	2000US-0246475.	CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703
PR	08-NOV-2000;	2000US-0246476.	CC	to AAK87694 represent human immune/hematopoietic antigen genomic
PR	08-NOV-2000;	2000US-0246477.	CC	sequences from the present invention. AAK54942 to AAK54950 and AAK82169
PR	08-NOV-2000;	2000US-0246478.	CC	represent sequences used in the exemplification of the present invention.
PR	08-NOV-2000;	2000US-0246523.	CC	
PR	08-NOV-2000;	2000US-0246524.	XX	
PR	08-NOV-2000;	2000US-0246525.	XX	Sequence 71 AA;
PR	08-NOV-2000;	2000US-0246526.	XX	
PR	08-NOV-2000;	2000US-0246527.	XX	Query Match 80.0%; Score 32; DB 22; Length 71;
PR	08-NOV-2000;	2000US-0246528.	XX	Best Local Similarity 75.0%; Pred. No. 48;
PR	08-NOV-2000;	2000US-0246532.	XX	Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
PR	08-NOV-2000;	2000US-0246532.	XX	
PR	08-NOV-2000;	2000US-0246609.	XX	
PR	08-NOV-2000;	2000US-0246610.	XX	
PR	08-NOV-2000;	2000US-0246611.	XX	
PR	08-NOV-2000;	2000US-0246613.	XX	
PR	17-NOV-2000;	2000US-0249207.	XX	
PR	17-NOV-2000;	2000US-0249208.	XX	
PR	17-NOV-2000;	2000US-0249209.	XX	
PR	17-NOV-2000;	2000US-0249210.	XX	
PR	17-NOV-2000;	2000US-0249211.	XX	
PR	17-NOV-2000;	2000US-0249212.	XX	ABG19347 standard; Protein; 73 AA.
PR	17-NOV-2000;	2000US-0249213.	XX	
PR	17-NOV-2000;	2000US-0249214.	XX	AC ABG19347;
PR	17-NOV-2000;	2000US-0249215.	XX	
PR	17-NOV-2000;	2000US-0249216.	XX	DT 18-FEB-2002 (first entry)
PR	17-NOV-2000;	2000US-0249217.	XX	
PR	17-NOV-2000;	2000US-0249218.	XX	Novel human diagnostic protein #19338.
PR	17-NOV-2000;	2000US-0249244.	XX	
PR	17-NOV-2000;	2000US-0249245.	XX	Human; chromosome mapping; gene mapping; gene therapy; forensic;
PR	17-NOV-2000;	2000US-0249264.	XX	food supplement; medical imaging; diagnostic; genetic disorder.
PR	17-NOV-2000;	2000US-0249265.	XX	
PR	17-NOV-2000;	2000US-0249297.	XX	Homo sapiens.
PR	17-NOV-2000;	2000US-0249299.	XX	WO200175067-A2.
PR	17-NOV-2000;	2000US-0249300.	XX	
PR	01-DEC-2000;	2000US-0250160.	XX	PD 11-OCT-2001.
PR	01-DEC-2000;	2000US-0250391.	XX	
PR	05-DEC-2000;	2000US-0251030.	XX	
PR	05-DEC-2000;	2000US-0251988.	XX	30-MAR-2001; 2001WO-US08631.
PR	05-DEC-2000;	2000US-0256719.	XX	
PR	06-DEC-2000;	2000US-0251479.	XX	31-MAR-2000; 2000US-0540217.
PR	08-DEC-2000;	2000US-0251856.	XX	23-AUG-2000; 2000US-0649167.
PR	08-DEC-2000;	2000US-0251868.	XX	(HYSE-) HYSEQ INC.
PR	08-DEC-2000;	2000US-0251869.	XX	
PR	08-DEC-2000;	2000US-0251989.	XX	Drmanac RT, Liu C, Tang YT;
PR	08-DEC-2000;	2000US-0251990.	XX	
PR	11-DEC-2000;	2000US-0254097.	XX	WPI: 2001-639362/73.
PR	05-JAN-2001;	2001US-0259678.	XX	N-PSDB; AAS83534.
XX			XX	
PA	(HUMA-) HUMAN GENOME SCI INC.		PT	New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 XX Claim 20; SEQ ID NO 49706; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 73 AA;

Query Match 80.0%; Score 32; DB 22; Length 73;
 Best Local Similarity 75.0%; Pred. No. 50;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9

Db 42 LLLPLQLM 49

RESULT 12

AAW22303
 ID AAW22303 standard; Protein; 267 AA.

XX AAW22303;

XX 24-NOV-1997 (first entry)

XX Rat CRTI.

XX Rat; CRTI; expression; liver cancer; tissue; antibody; probe;
 KW detection; assay; diagnosis.

XX Rattus rattus.

XX WO9710333-A1.

XX 20-MAR-1997.

XX 17-SEP-1996; 96WO-JP02654.

XX 30-AUG-1996; 96JP-0229469.

XX 14-SEP-1995; 95JP-0236264.

XX 27-NOV-1995; 95JP-0331023.

XX 20-JUN-1996; 96JP-0179885.

XX (SUME) SUMITOMO ELECTRIC IND CO.

XX Kishimoto T, Kokura K, Kumagai Y, Makino Y, Tamura T;

XX WPI; 1997-202229/18.

XX N-PSDB; AAT73337.

XX Proteins with elevated expression in liver cancer, and related DNA -

PT for production of antibodies useful in the diagnosis and monitoring
 PT of liver cancer
 XX
 PS Claim 1; Pages 69-70; 119pp; Japanese.
 XX The present sequence is rat CRTI, which has an elevated
 CC expression in liver cancer tissue. Anti-CRTI antibodies or CRTI
 CC cDNA can be used to detect or assay for CRTI or CRTI mRNA,
 CC indicating in which tissues CRTI expression is elevated, and
 CC therefore allowing liver cancer to be diagnosed and monitored.

XX Sequence 267 AA;

Query Match 80.0%; Score 32; DB 18; Length 267;

Best Local Similarity 100.0%; Pred. No. 1.9e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 LLPLQL 9

Db 188 LLPLQL 194

RESULT 13

AAB38062

ID AAB38062 standard; Peptide; 46 AA.

XX AAB38062;

XX 31-JAN-2001 (first entry)

XX Fragment of human secreted protein encoded by gene 18 clone HKAFH74.

XX Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
 KW anti-allergic; hepatotropic; antidiabetic; antinflammatory; antitumor;
 KW vulnary; anticonvulsant; antibacterial; antifungal; antiparasitic;
 KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; human; secreted protein.

XX Homo sapiens.

XX WO200055371-A1.

XX 21-SEP-2000.

XX 16-MAR-2000; 2000WO-US06783.

XX 18-MAR-1999; 99US-0125055.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Ruben SM, Nil J, Ebner R, Rosen CA, Shi Y, Birse C, Florence K;

PI Komatsoulis G, Lafleur DW, Moore PA, Olsen HS, Young PE;

XX WPI; 2000-594448/56.

XX New nucleic acid molecules encoding 27 human secreted proteins for
 PT diagnosing, preventing, treating or ameliorating medical conditions and
 PT used as food additives or preservatives -

XX Disclosure; Page 41; 453pp; English.

XX Sequences AAB37984-B38019 represent the amino acid sequences of 27
 CC human secreted proteins encoded by the genes AAC69084-C69119. The genes
 CC and proteins are useful for preventing, ameliorating or treating medical
 CC conditions, e.g. by protein or gene therapy. The genes are isolated from
 CC a range of human tissues disclosed in the specification. The nucleic
 CC acids, proteins, antibodies and (ant)agonists are useful in the
 CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
 CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
 CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative

CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 46 AA;
Query Match 77.5%; Score 31; DB 21; Length 46;
Best Local Similarity 77.8%; Pred. No. 48;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 SLLPLQLIL 9
Db 31 SLLPLLLL 39
RESULT 14
AAM79694
ID AAM79694 standard; Protein; 56 AA.
XX
AC AAM79694;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 3340.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI: 2001-476283/51.
DR N-PSDB; AAK52827.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 308; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.

CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 56 AA;
Query Match 77.5%; Score 31; DB 22; Length 56;
Best Local Similarity 85.7%; Pred. No. 59;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 3 LLPLQLIL 9
Db 1 LLPLQVL 7
RESULT 15
ABG12769
ID ABG12769 standard; Protein; 89 AA.
XX
AC ABG12769;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #12760.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS76956.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 43128; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX
SQ Sequence 89 AA;
Query Match 77.5%; Score 31; DB 22; Length 89;
Best Local Similarity 55.6%; Pred. No. 95;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;
QY 1 SLLPLQL 9
:|::|||:|
Db 17 NLIPLQLL 25

RESULT 16
ABG13602
ID ABG13602 standard; Protein; 89 AA.
XX
AC ABG13602;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #13593.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX

DR WPI; 2001-639362/73.
DR N-PSDB; AAS77789.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX

PS Claim 20; SEQ ID No 43961; 103pp; English.

XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 89 AA;

Query Match 77.5%; Score 31; DB 22; Length 89;
Best Local Similarity 55.6%; Pred. No. 95;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQL 9
:|::|||:|
Db 17 NLIPLQLL 25

RESULT 17
AAB63580
ID AAB63580 standard; Protein; 142 AA.
XX
AC AAB63580;
XX
DT 26-MAR-2001 (first entry)
XX
DE Human gastric cancer associated antigen protein sequence SEQ ID NO:942.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
XX
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
PD 07-DEC-2000.
XX
PF 26-MAY-2000; 2000WO-US14749.
XX
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
XX
PI Obata Y;
XX
DR WPI; 2001-025274/03.
XX

PT Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX

PS Example 1; Page 625; 799pp; English.

XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP), respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX

SQ Sequence 142 AA;

Query Match 77.5%; Score 31; DB 22; Length 142;
Best Local Similarity 75.0%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLLPLQL 9
:|::|||:|
Db 47 VLLPLQLL 54

RESULT 18
AAB63578
ID AAB63578 standard; Protein; 156 AA.

```
XX AAB63578;
AC
XX
DT 26-MAR-2001 (first entry)
DE Human gastric cancer associated antigen protein sequence SEQ ID NO:940.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
OS Homo sapiens.
XX
PN WO200073801-A2.
XX
XX 07-DEC-2000.
XX
XX 26-MAY-2000; 2000WO-US14749.
PR 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PI Obata Y;
XX
DR WPI; 2001-025274/03.
XX
XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
XX Example 1; Page 624; 799pp; English.
XX
CC AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
CC represent nucleotide sequences encoding human breast, gastric and
CC prostate cancer associated antigen precursors (CAAP) respectively.
CC AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
CC represent human breast, gastric and prostate CAAP protein sequence
CC respectively. CAAPs have cytostatic activity and can be used in the
CC production of cancer vaccines. The human CAAP proteins, peptides, nucleic
CC acids or anti-CAAP antibodies are useful for diagnosing and treating a
CC condition characterised by expression of an abnormal amount of a protein,
CC e.g. cancer.
XX
XX Sequence 156 AA;
XX
XX Query Match 77.5%; Score 31; DB 22; Length 156;
XX Best Local Similarity 75.0%; Pred. No. 1.7e+02;
XX Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
QY 2 LLLPLQLL 9
DB 47 VLLPLQLL 54
:|||||:
:|||||:

RESULT 19
AAB63582
ID AAB63582 standard; Protein; 159 AA.
XX
AC AAB63582;
XX
XX 26-MAR-2001 (first entry)
XX
DE Human gastric cancer associated antigen protein sequence SEQ ID NO:944.
XX
KW Human; breast cancer; gastric cancer; prostate cancer; diagnosis;
KW cancer associated antigen; cytostatic; cancer vaccine.
OS Homo sapiens.
XX
XX WO200073801-A2.
XX
XX
```

```
PD 07-DEC-2000.
XX
XX 26-MAY-2000; 2000WO-US14749.
XX
XX 28-MAY-1999; 99US-0136526.
PR 10-SEP-1999; 99US-0153454.
XX
PA (LUDW-) LUDWIG INST CANCER RES.
PI Obata Y;
XX
XX WPI; 2001-025274/03.
XX
XX Nucleic acids encoding breast, gastric and prostate cancer associated
PT antigen precursors, useful for diagnosing and treating a condition
PT characterized by expression of an abnormal amount of a protein, e.g.
PT cancer -
XX
XX Example 1; Page 625-626; 799pp; English.
XX
XX AAF22422 to AAF22626, AAF22627 to AAF22773 and AAF22774 to AAF23014
XX represent nucleotide sequences encoding human breast, gastric and
XX prostate cancer associated antigen precursors (CAAP) respectively.
XX AAB63232 to AAB63467, AAB63468 to AAB63721 and AAB63722 to AAB63970
XX represent human breast, gastric and prostate CAAP protein sequence
XX respectively. CAAPs have cytostatic activity and can be used in the
XX production of cancer vaccines. The human CAAP proteins, peptides, nucleic
XX acids or anti-CAAP antibodies are useful for diagnosing and treating a
XX condition characterised by expression of an abnormal amount of a protein,
XX e.g. cancer.
XX
XX Sequence 159 AA;
XX
XX Query Match 77.5%; Score 31; DB 22; Length 159;
XX Best Local Similarity 75.0%; Pred. No. 1.7e+02;
XX Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
XX
QY 2 LLLPLQLL 9
DB 47 VLLPLQLL 54
:|||||:
:|||||:

RESULT 20
AAB71026
ID AAB71026 standard; Protein; 189 AA.
XX
AC AAB71026;
XX
XX 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 39870.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX
XX Drosophila melanogaster.
XX
XX WO200171042-A2.
XX
XX 27-SEP-2001.
XX
XX 23-MAR-2001; 2001WO-US09231.
XX
XX 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE ) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
XX WPI; 2001-656860/75.
XX N-PSDB; ABL15129.
XX
```

PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
XX interactions -
PS Disclosure; SEQ ID NO 39870; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABU16176-ABU30511), expressed DNA
CC sequences (ABU01840-ABU16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 189 AA;
Query Match 77.5%; Score 31; DB 22; Length 189;
Best Local Similarity 87.5%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LLLPLQL 9
Db 42 LLLPLNIL 49
RESULT 21
AAB38058
ID AAB38058 standard; Peptide; 211 AA.
XX
AC AAB38058;
XX
DT 31-JAN-2001 (first entry)
XX
DE Fragment of human secreted protein encoded by gene 18 clone HKAHF74.
XX
KW Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW neurological disease; infection; human; secreted protein.
XX
OS Homo sapiens.
XX
XX WO20005371-A1.
XX
PD 21-SEP-2000.
XX
XX 16-MAR-2000; 2000WO-US06783.
XX
PR 18-MAR-1999; 99US-0125055.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Ruben SM, Ni J, Ebner R, Rosen CA, Shi Y, Birse C, Florence K;
PI Komatsoulis G, Lafleur DW, Moore PA, Olsen HS, Young PE;
XX
XX WPI; 2000-594448/56.
XX
XX New nucleic acid molecules encoding 27 human secreted proteins for
PT diagnosing, preventing, treating or ameliorating medical conditions and
PT used as food additives or preservatives -
XX
XX Disclosure; Page 41; 453pp; English.
XX
CC Sequences AAB37984-B38019 represent the amino acid sequences of 27
CC human secreted proteins encoded by the genes AAC69084-C69119. The genes
CC and proteins are useful for preventing, ameliorating or treating medical
CC conditions, e.g. by protein or gene therapy. The genes are isolated from
CC a range of human tissues disclosed in the specification. The nucleic

CC acids, proteins, antibodies and (ant)agonists are useful in the
CC diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
CC ovarian cancer, and other cancers of the adrenal gland, bone, bone
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus,
CC Crohn's disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC and parasitic infections.
XX
SQ Sequence 211 AA;
Query Match 77.5%; Score 31; DB 21; Length 211;
Best Local Similarity 77.8%; Pred. No. 2.3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 SLLPLQL 9
Db 162 SLLPLLLL 170
RESULT 22
AAB24037
ID AAB24037 standard; Protein; 246 AA.
XX
AC AAB24037;
XX
DT 25-JAN-2001 (first entry)
XX
DE Human PRO1555 protein sequence SEQ ID NO:49.
XX
KW Human; tumour; diagnosis; neoplastic disease; proliferation; cancer;
KW identification; tumorigenesis; anticancer; detection.
XX
OS Homo sapiens.
XX
XX WO200053750-A1.
XX
PD 14-SEP-2000.
XX
PF 02-DEC-1999; 99WO-US28551.
XX
PR 08-MAR-1999; 99WO-US05028.
PR 01-SEP-1999; 99WO-US20111.
PR 29-OCT-1999; 99US-0162506.
PR 30-NOV-1999; 99WO-US28313.
PR 01-DEC-1999; 99WO-US28634.
XX
XX (GETH) GENENTECH INC.
XX
XX Botstein D, Goddard A, Gurney AL, Roy MA, Watanabe CK, Wood WI;
PI N-PSDB; AAC58119.
XX
XX WPI; 2000-594320/56.
XX
XX Antibodies specific for PRO polypeptides, used to diagnose and inhibit
PT the growth of tumors in mammals, and to identify inhibitors of PRO
PT polypeptide activity or expression -
XX
XX Claim 61; Fig 34; 226pp; English.
XX
CC The present invention describes an antibody that binds to a human
CC protein (I) selected from: PRO381; PRO1269; PRO1410; PRO1755; PRO1780;
CC PRO3434; PRO1927; PRO3567; PRO1293; PRO1303; PRO4344; PRO4354;
CC PRO4397; PRO4407; PRO1555; PRO1096; PRO2262. (I) has
CC anticancer activity and can be used to diagnose tumours in mammals, by
CC detecting complex formation when the antibody is contacted with test
CC cells. Increased expression of genes encoding (I) can also be detected
CC to diagnose tumours. Agents which inhibit the activity of (I),
CC especially the antibodies, or an antisense oligonucleotide which
CC hybridises to genes encoding (I), can be used to inhibit tumour growth,

CC preferably by inducing cell death. Methods from the present invention
CC can be used to identify compounds which inhibit the biological activity
CC of (1). AAC58019 to AAC58102 represent PCR primers and hybridisation
CC probes used in examples from the present invention for human PRO
CC sequences. AAC58103 to AAC58122 and AAB24021 to AAB24040 represent human
CC PRO polynucleotide and protein sequences given in the exemplification of
CC the present invention.

XX
SQ Sequence 246 AA;

Query Match 77.5%; Score 31; DB 21; Length 246;
Best Local Similarity 77.8%; Pred. No. 2.7e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLPLQLIL 9
 |||||:1
Db 197 SLLPLQLLL 205

RESULT 23

AAY99438
ID AAY99438 standard; Protein; 246 AA.

XX AC AAY99438;

XX DT 08-AUG-2000 (first entry)

XX DE Human PRO1555 (UNQ763) amino acid sequence SEQ ID NO:338.

XX KW Human; PRO polypeptide; membrane bound protein; receptor; diagnosis;
KW transmembrane; secretion; immunoadhesion; pharmaceutical; screening.

XX OS Homo sapiens.

XX PN WO200012708-A2.

XX PD 09-MAR-2000.

XX PF 01-SEP-1999; 99WO-US20111.

XX PR 01-SEP-1998; 98US-0098716.

PR 01-SEP-1998; 98US-0098749.

PR 01-SEP-1998; 98US-0098750.

PR 02-SEP-1998; 98US-0098803.

PR 02-SEP-1998; 98US-0098821.

PR 02-SEP-1998; 98US-0098843.

PR 09-SEP-1998; 98US-0099536.

PR 09-SEP-1998; 98US-0099596.

PR 09-SEP-1998; 98US-0099598.

PR 09-SEP-1998; 98US-0099602.

PR 09-SEP-1998; 98US-0099642.

PR 10-SEP-1998; 98US-0099741.

PR 10-SEP-1998; 98US-0099754.

PR 10-SEP-1998; 98US-0099763.

PR 10-SEP-1998; 98US-0099792.

PR 10-SEP-1998; 98US-0099808.

PR 10-SEP-1998; 98US-0099812.

PR 10-SEP-1998; 98US-0099815.

PR 10-SEP-1998; 98US-0099816.

PR 15-SEP-1998; 98US-0100385.

PR 15-SEP-1998; 98US-0100388.

PR 15-SEP-1998; 98US-0100390.

PR 16-SEP-1998; 98US-0100584.

PR 16-SEP-1998; 98US-0100627.

PR 16-SEP-1998; 98US-0100661.

PR 16-SEP-1998; 98US-0100662.

PR 16-SEP-1998; 98US-0100664.

PR 17-SEP-1998; 98US-0100663.

PR 17-SEP-1998; 98US-0100684.

PR 17-SEP-1998; 98US-0100710.

PR 17-SEP-1998; 98US-0100711.

PR 17-SEP-1998; 98US-0100919.

PR 17-SEP-1998; 98US-0100930.

PR 18-SEP-1998; 98US-0100848.
PR 18-SEP-1998; 98US-0100849.
PR 18-SEP-1998; 98US-0101014.
PR 18-SEP-1998; 98US-0101068.
PR 18-SEP-1998; 98US-0101071.
PR 22-SEP-1998; 98US-0101279.
PR 23-SEP-1998; 98US-0101471.
PR 23-SEP-1998; 98US-0101472.
PR 23-SEP-1998; 98US-0101474.
PR 23-SEP-1998; 98US-0101475.
PR 23-SEP-1998; 98US-0101476.
PR 23-SEP-1998; 98US-0101477.
PR 23-SEP-1998; 98US-0101479.
PR 24-SEP-1998; 98US-0101738.
PR 24-SEP-1998; 98US-0101741.
PR 24-SEP-1998; 98US-0101743.
PR 24-SEP-1998; 98US-0101915.
PR 24-SEP-1998; 98US-0101916.
PR 29-SEP-1998; 98US-0102207.
PR 29-SEP-1998; 98US-0102240.
PR 29-SEP-1998; 98US-0102307.
PR 29-SEP-1998; 98US-0102330.
PR 29-SEP-1998; 98US-0102331.
PR 30-SEP-1998; 98US-0102484.
PR 30-SEP-1998; 98US-0102570.
PR 30-SEP-1998; 98US-0102571.
PR 01-OCT-1998; 98US-0102584.
PR 01-OCT-1998; 98US-0102687.
PR 02-OCT-1998; 98US-0102965.
PR 06-OCT-1998; 98US-0103258.
PR 06-OCT-1998; 98US-0103449.
PR 07-OCT-1998; 98US-0103314.
PR 07-OCT-1998; 98US-0103315.
PR 07-OCT-1998; 98US-0103328.
PR 07-OCT-1998; 98US-0103395.
PR 07-OCT-1998; 98US-0103396.
PR 07-OCT-1998; 98US-0103401.
PR 08-OCT-1998; 98US-0103633.
PR 08-OCT-1998; 98US-0103678.
PR 08-OCT-1998; 98US-0103679.
PR 08-OCT-1998; 98US-0103711.
PR 20-OCT-1998; 98US-0104257.
PR 20-OCT-1998; 98US-0104987.
PR 20-OCT-1998; 98US-0105000.
PR 21-OCT-1998; 98US-0105002.
PR 22-OCT-1998; 98US-0105104.
PR 22-OCT-1998; 98US-0105169.
PR 22-OCT-1998; 98US-0105266.
PR 26-OCT-1998; 98US-0105693.
PR 26-OCT-1998; 98US-0105694.
PR 27-OCT-1998; 98US-0105807.
PR 27-OCT-1998; 98US-0105881.
PR 27-OCT-1998; 98US-0105882.
PR 27-OCT-1998; 98US-0106062.
PR 28-OCT-1998; 98US-0106023.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106029.
PR 28-OCT-1998; 98US-0106030.
PR 28-OCT-1998; 98US-0106032.
PR 28-OCT-1998; 98US-0106033.
PR 28-OCT-1998; 98US-0106178.
PR 29-OCT-1998; 98US-0106248.
PR 29-OCT-1998; 98US-0106384.
PR 29-OCT-1998; 98US-0106500.
PR 30-OCT-1998; 98US-0106464.
PR 03-NOV-1998; 98US-0106856.
PR 03-NOV-1998; 98US-0106902.
PR 03-NOV-1998; 98US-0106905.
PR 03-NOV-1998; 98US-0106919.
PR 03-NOV-1998; 98US-0106932.
PR 03-NOV-1998; 98US-0106934.
PR 17-NOV-1998; 98US-0107783.
PR 17-NOV-1998; 98US-0108775.

PR 17-NOV-1998; 98US-0108779.
 PR 17-NOV-1998; 98US-0108787.
 PR 17-NOV-1998; 98US-0108788.
 PR 17-NOV-1998; 98US-0108801.
 PR 17-NOV-1998; 98US-0108802.
 PR 17-NOV-1998; 98US-0108806.
 PR 17-NOV-1998; 98US-0108807.
 PR 17-NOV-1998; 98US-0108867.
 PR 17-NOV-1998; 98US-0108925.
 PR 18-NOV-1998; 98US-0108848.
 PR 18-NOV-1998; 98US-0108849.
 PR 18-NOV-1998; 98US-0108850.
 PR 18-NOV-1998; 98US-0108851.
 PR 18-NOV-1998; 98US-0108852.
 PR 18-NOV-1998; 98US-0108858.
 PR 18-NOV-1998; 98US-0108904.
 XX
 PA (GETH) GENENTECH INC.

XX Baker K, Goddard A, Gurney AL, Smith V, Watanabe CK, Wood WL;

XX PI 2000-237871/20.

XX WPI; 2000-237871/20.

XX N-PSDB; AAA37120.

XX New mammalian DNA sequences encoding transmembrane, receptor or
 PT secreted PRO polypeptides, useful for screening of potential peptide or
 PT small molecule inhibitors of the relevant receptor/ligand interactions

XX Claim 12; Fig 198; 773pp; English.

XX AAA37022 to AAA37144 encode the new isolated human transmembrane,
 CC receptor or secreted PRO polypeptides given in AAY9340 to AAY99462. The
 CC transmembrane and receptor PRO proteins can be used for screening of
 CC potential peptide or small molecule inhibitors of the relevant
 CC receptor/ligand interactions. The polypeptides and nucleotide sequences
 CC encoding them have various industrial applications, including uses as
 CC pharmaceutical and diagnostic agents. AAA37145 to AAA37330 represent
 CC PCR primers and hybridisation probes used in the isolation of the PRO
 CC polypeptides from the present invention.

XX SQ Sequence 246 AA;

Query Match 77.5%; Score 31; DB 21; Length 246;

Best Local Similarity 77.8%; Pred. No. 2.7e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SLLPLQIL 9

||||| :|

DB 197 SLLPLLLL 205

RESULT 24

AA866187

ID AAB66187 standard; protein; 246 AA.

XX AC AAB66187;

XX DT 02-APR-2001 (first entry)

XX Protein of the invention #99.

XX DE Secreted; transmembrane; gene therapy.

XX OS Unidentified.

XX PN WO200078961-A1.

XX XX 28-DEC-2000.

XX PF 18-FEB-2000; 2000WO-US04342.

XX PR 23-JUN-1999; 99US-0141037.

XX PR 20-JUL-1999; 99US-0144758.

PR 26-JUL-1999; 99US-0145698.
 PR 01-SEP-1999; 99WO-US20111.
 PR 29-OCT-1999; 99US-0162506.
 PR 30-NOV-1999; 99WO-US28313.
 PR 02-DEC-1999; 99WO-US28551.
 PR 16-DEC-1999; 99WO-US30095.
 PR 05-JAN-2000; 2000WO-US00219.
 PR 06-JAN-2000; 2000WO-US00376.
 XX
 PA (GETH) GENENTECH INC.

XX Baker KP, Botstein D, Desnoyers L, Eaton DL, Ferrara N, Fong S;

XX PI Cao W, Goddard A, Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ;

XX PI Pan J, Paoni NF, Roy MA, Smith V, Stewart TA, Tumas D;

XX PI Watanabe CK, Williams PW, Wood WL;

XX WPI; 2001-071395/08.

XX Secreted and transmembrane proteins and nucleic acids designated PRO.
 PT useful as hybridization probes, in chromosome and gene mapping and gene
 PT therapy -

XX Claim 1; Fig 198; 787pp; English.

XX The present invention relates to secreted and transmembrane proteins.
 CC These proteins and the DNA encoding them may be used as hybridization
 CC probes, in chromosome and gene mapping and in the generation of
 CC anti-sense RNA and DNA. They may also be used to generate either
 CC transgenic animals or knockout animals which are in turn useful for
 CC development and screening of therapeutically useful reagents.
 CC The nucleic acids may also be used in gene therapy.

XX SQ Sequence 246 AA;

Query Match 77.5%; Score 31; DB 22; Length 246;

Best Local Similarity 77.8%; Pred. No. 2.7e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 SLLPLQIL 9

||||| :|

DB 197 SLLPLLLL 205

RESULT 25

AAU08753

ID AAU08753 standard; Protein; 278 AA.

XX AC AAU08753;

XX DT 03-JAN-2002 (first entry)

XX Human insulin-like growth factor binding protein-like polypeptide #3.

XX Insulin-like growth factor binding protein; IGFBP; human; cancer;
 KW female reproduction; embryo development; food supplement; gene mapping;
 KW medical imaging; autoimmune disease; nervous system disease; cytostatic;
 KW cerebrovascular disease; wound healing; gynaecological; antifertility;
 KW gene therapy; vulnery.

XX Homo sapiens.

XX OS Location/Qualifiers

XX Key Peptide 1..27

FT /note= "Signal peptide"

FT Protein 28..278

FT /note= "Mature human IGFBP-like polypeptide #3"

FT Domain 61..76

FT /note= "IGFBP signature"

XX WO200175064-A2.

XX PN 11-OCT-2001.

XX PD

XX XX

PF 30-MAR-2001; 2001WO-US10462.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
PR 14-FEB-2001; 2001US-0784748.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Yamazaki V, Asundi V, Drmanac RT, Liu C, Tang YT;
XX WPI; 2001-626426/72.
DR N-PSDB; AAS14771.
XX
XX New insulin-like growth factor binding protein-like polypeptide and
PT encoding polynucleotides, useful for treating cancer, infertility, and
PT arthritis, and for increasing wound healing
XX Claim 9; Page 123-124; 130pp; English.
XX
XX The invention relates to isolated insulin-like growth factor binding
CC protein-like (IGFBP-like) polypeptides and their associated
CC polynucleotides. The DNA sequences can be detected by contacting a sample
CC with nucleic acid primers that anneal to the DNA and amplifying a product
CC comprising a portion of the sequence. Detection of the product indicates
CC the presence of DNA. The protein sequences can be detected by contacting
CC a sample with a compound that binds to the polypeptide to form a complex.
CC Detection of the complex indicates the presence of the protein. The
CC sequences of the invention are useful for treating a subject having a
CC need to inhibit activity or expression of IGFBP-like sequences. This
CC involves administering an antagonist of the polypeptide, a polynucleotide
CC that inhibits the expression of the nucleotide sequence or a therapeutic
CC amount of the polypeptide that competes for its ligand and a carrier. The
CC sequences are useful in treatment of disorders such as cancer, or to
CC promote female reproductive health and embryo development. They can also
CC be used in food supplements, in medical imaging and in gene mapping. The
CC sequences can be used in the treatment and prevention of autoimmune
CC diseases, nervous system diseases, cerebrovascular diseases and
CC infertility and for enhancing wound healing. This sequence represents a
CC human IGFBP-like polypeptide.
XX
XX Sequence 278 AA;
SQ

Query Match 77.5%; Score 31; DB 22; Length 278;
Best Local Similarity 77.8%; Pred. No. 3e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLPLQL 9
Db 5 SLLPLQLL 13
||||| :|

RESULT 26
ABB62737
ID ABB62737 standard; Protein; 316 AA.
XX
XX ABB62737;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 15003.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
KW
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR

PR 11-JUL-2000; 2000US-0614150.
XX
XX (PEKE) PE CORP NY.
PA
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR N-PSDB; ABL06840.
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX Disclosure; SEQ ID NO 15003; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 316 AA;
SQ

Query Match 77.5%; Score 31; DB 22; Length 316;
Best Local Similarity 66.7%; Pred. No. 3.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 SLLPLQL 9
Db 238 SLLPLQV 246
||||| :|

RESULT 27
ABG01683
ID ABG01683 standard; Protein; 390 AA.
XX
XX ABG01683;
AC
XX
XX 13-FEB-2002 (first entry)
DT
XX
XX Novel human diagnostic protein #1674.
DE
XX
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
KW
XX
XX Homo sapiens.
OS
XX
XX WO200175067-A2.
PN
XX
XX 11-OCT-2001.
PD
XX
XX 30-MAR-2001; 2001WO-US08631.
PF
XX
XX 31-MAR-2000; 2000US-0540217.
PR
XX
XX 23-AUG-2000; 2000US-0649167.
PR
XX
XX (HYSE-) HYSEQ INC.
PA
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS65870.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
PT

XX PS Claim 20: SEQ ID No 32042; 103pp; English.

XX CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations in

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX CC

SQ Sequence 390 AA;

Query Match 77.5%; Score 31; DB 22; Length 390;

Best Local Similarity 77.8%; Pred. No. 4.3e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

Db 5 SLLPLLLL 13

IIIIII :I

RESULT 28

AAU08755

ID AAU08755 standard; Protein; 390 AA.

XX CC

AC AAU08755;

XX CC

DT 03-JAN-2002 (first entry)

XX CC

DE Human insulin-like growth factor binding protein-like polypeptide #2.

XX CC

KW Insulin-like growth factor binding protein; IGFBP; human; cancer;

KW female reproduction; embryo development; food supplement; gene mapping;

KW medical imaging; autoimmune disease; nervous system disease; cytostatic;

KW cerebrovascular disease; wound healing; gynaecological; antiinfertility;

KW gene therapy; vulnerary.

XX CC

OS Homo sapiens.

XX CC

PN WO200175064-A2.

XX CC

PD 11-OCT-2001.

XX CC

PF 30-MAR-2001; 2001WO-US10462.

XX CC

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PR 14-FEB-2001; 2001US-0784748.

XX CC

PA (HYSE-) HYSEQ INC.

XX CC

PI Yamazaki V, Asundi V, Drmanac RT, Liu C, Tang YT;

XX CC

XX WPI: 2001-626426/72.

DR N-PSDB; AAS14770.

DR CC

XX New insulin-like growth factor binding protein-like polypeptide and

PT encoding polynucleotides, useful for treating cancer, infertility, and

PT arthritis, and for increasing wound healing -

XX CC

PS Claim 9; Page 108; 130pp; English.

XX CC

CC The invention relates to isolated insulin-like growth factor binding

CC protein-like (IGFBP-like) polypeptides and their associated

CC polynucleotides. The DNA sequences can be detected by contacting a sample

CC with nucleic acid primers that anneal to the DNA and amplifying a product

CC comprising a portion of the sequence. Detection of the product indicates

CC the presence of DNA. The protein sequences can be detected by contacting

CC a sample with a compound that binds to the polypeptide to form a complex.

CC Detection of the complex indicates the presence of the protein. The

CC sequences of the invention are useful for treating a subject having a

CC need to inhibit activity or expression of IGFBP-like sequences. This

CC involves administering an antagonist of the polypeptide, a polynucleotide

CC that inhibits the expression of the nucleotide sequence or a therapeutic

CC amount of the polypeptide that competes for its ligand and a carrier. The

CC sequences are useful in treatment of disorders such as cancer, or to

CC promote female reproductive health and embryo development. They can also

CC be used in food supplements, in medical imaging and in gene mapping. The

CC sequences can be used in the treatment and prevention of autoimmune

CC diseases, nervous system diseases, cerebrovascular diseases and

CC infertility and for enhancing wound healing. This sequence represents a

CC human IGFBP-like polypeptide.

XX CC

SQ Sequence 390 AA;

Query Match 77.5%; Score 31; DB 22; Length 390;

Best Local Similarity 77.8%; Pred. No. 4.3e+02;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 SLLPLQL 9

Db 5 SLLPLLLL 13

IIIIII :I

RESULT 29

AAU17125

ID AAU17125 standard; Protein; 393 AA.

XX CC

AC AAU17125;

XX CC

DT 07-NOV-2001 (first entry)

XX CC

DE Novel signal transduction pathway protein, Seq ID 690.

XX CC

KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;

KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;

KW immune system disorder; rheumatoid arthritis; inflammatory condition;

KW organ transplant rejection; infection; hepatitis C; blood disorder;

KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;

KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;

KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;

KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;

KW reproductive system; gastrointestinal; liver disorder; AIDS;

KW acquired immune deficiency syndrome.

XX CC

OS Homo sapiens.

XX CC

PN WO200154733-A1.

XX CC

PD 02-AUG-2001.

XX CC

PF 17-JAN-2001; 2001WO-US01312.

XX CC

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209457.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231143.
PR 08-SEP-2000; 2000US-0231141.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249266.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PR (HUMA-) HUMAN GENOME SCI INC.
PR Rosen CA, Barash SC, Ruben SM;
PR WPI; 2001-465460/50.
PR N-PSDB; AAS27042.
PR Novel polypeptides useful for diagnosing, treating, preventing and/or
PR prognosing disorders related to the proteins, including cancers, immune
PR disorders and neuronal disorders -
PR Claim 1; SEQ ID No 690; 880pp; English.
PR The invention relates to novel isolated polypeptides (I), and
PR polynucleotides (II). (I), (II) and the antibody to (I) are useful for
PR diagnosing, preventing and treating diseases including immune system
PR disorders (e.g. congenital and acquired immunodeficiencies, autoimmune

disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ transplant rejections and graft versus host disease, infectious diseases (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and other blood-related disorders (sickle cell anemia), myeloproliferative disorders, primary haematopoietic disorders, hyperproliferative disorders (e.g. Gaucher's disease and cancer), neurodegenerative disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal disorders (e.g. glomerulonephritis), cardiovascular disorders (e.g. arrhythmia), respiratory disorders, dermatological disorders, wound healing, epithelial cell proliferation, endocrine disorders (e.g. Addison's disease), reproductive system disorders, gastrointestinal disorder (inflammatory disorders), liver disorders (cirrhosis), as stimulators of B-cell responsiveness to pathogens, activators of T-cells, to induce higher affinity antibodies, and as a means to induce tumour proliferation in pathologies e.g. acquired immune deficiency syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction pathway protein, amino acid sequences of the invention.

Query Match 77.5%; Score 31; DB 22; Length 393;
Best Local Similarity 66.7%; Pred. No. 4.3e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 SLLLPLOQL 9
:|||||:
Db 172 ALLPLREL 180

RESULT 30
AAU23643
ID AAU23643 standard; Protein: 585 AA.
XX
AC AAU23643;
XX
DT 06-SEP-1999 (first entry)
XX
DE Protein encoded by a reading frame of the porcine circovirus genome.
XX
KW MAP; piglet fatal wasting disease; vaccine; circovirus infection;
KW gene therapy.
XX
OS Porcine circovirus.
XX
EH Key Location/Qualifiers
FT Misc-difference 1..585
/note= "all X residues are encoded by stop codons"
XX
PN FR2772047-A1.
XX
PD 11-JUN-1999.
XX
PF 05-DEC-1997; 97FR-0015396.
XX
PR 05-DEC-1997; 97FR-0015396.
XX
PA (NAVE-) CENT NAT ETUD VETERINAIRES & ALIMENTAIRE.

PI Albina E, Arnould C, Blanchard P, Hutet E, Jestin A;
PI Le Cann P;
DR WPI: 1999-360000/31.
DR N-PSDB; AAX85593.
XX
PT Nucleotide sequence of porcine circovirus MAP - useful in vaccines
PT against MAP circovirus infection and in gene therapy
XX
PS Claim 7; Fig 3; 89pp; French.

XX The present sequence is encoded by a reading frame of the negative
CC strand of a porcine circovirus genome which is associated with MAP.
CC MAP is the french acronym for piglet fatal wasting disease. The
CC polypeptides can be used to detect anti-MAP antibodies. The antibodies

CC can be used to detect MAP antigens. The nucleotide sequences can be used
CC as probes or primers for detecting MAP nucleic acids. The nucleotide
CC sequences, polypeptides, vectors, (pseudo)viral particles, transformed
CC cells and compounds selected by the screening assay can be used in
CC pharmaceutical compositions. The polypeptides, nucleotide sequences,
CC vectors and transformed cells can be used in vaccines against MAP
CC circovirus infection. The vectors, (pseudo)viral particles and
XX transformed cells can be used for gene therapy.

SQ Sequence 585 AA;

Query Match 77.5%; Score 31; DB 20; Length 585;
Best Local Similarity 66.7%; Pred. No. 6.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLLPLOQL 9
:|||||:
Db 165 AYLPLQVL 173

RESULT 31
ABG25713
ID ABG25713 standard; Protein: 770 AA.
XX
AC ABG25713;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #25704.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

DR N-PSDB; AAS89900.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity

PS Claim 20; SEQ ID No 56072; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 770 AA;

Query Match 77.5%; Score 31; DB 22; Length 770;
 Best Local Similarity 87.5%; Pred. No. 8.6e+02;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 LLLPLQL 9
 II IIIII
 Db 327 LLAPLQL 334

RESULT 32
 AAG92740
 ID AAG92740 standard; Protein; 925 AA.
 AC AAG92740;

DT 26-SEP-2001 (first entry)
 XX C glutamicum protein fragment SEQ ID NO: 6494.
 DE
 XX

KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
 KW organic acid synthesis.

XX Corynebacterium glutamicum.

OS EPI108790-A2.

PN 20-JUN-2001.

PD 18-DEC-2000; 2000EP-0127688.

PF 16-DEC-1999; 99JP-0377484.

PR 07-APR-2000; 2000JP-0159162.

PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW) KYOWA HAKKO KOGYO KK.

PA Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
 XX Tateishi N, Senoh A, Ikeda M, Ozaki A;

PI WPI: 2001-376931/40.

DR N-PSDB; AAH67959.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying
 XX mutation point of a gene, measuring expression of a gene, analysing
 XX expression profile or pattern of a gene and identifying homologous gene

PS Claim 17: SEQ ID NO: 6494; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein
 XX sequences from the Coryneform bacterium Corynebacterium glutamicum. These
 XX are useful for identifying the mutation point of a gene derived from a
 XX mutant of coryneform bacterium, measuring expression amount and
 XX analysing the expression profile or expression pattern of a gene derived
 XX from Coryneform bacterium, and identifying a homologue of a gene derived
 XX from coryneform bacterium. Coryneform bacteria are useful for producing
 XX amino acids, nucleic acids, vitamins, saccharides and organic acids,
 XX particularly L-lysine. The present sequence is a protein described
 XX in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC European Patent Office.

XX Sequence 925 AA;

Query Match

Best Local Similarity 77.5%; Score 31; DB 22; Length 925;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 SLLPLQL 9
 III:III II
 Db 694 SLLPLLLIL 702

RESULT 33

ABG01394

ID ABG01394 standard; Protein; 1062 AA.

XX AC ABG01394;

XX DT 13-FEB-2002 (first entry)

XX Novel human diagnostic protein #1385.

DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

DR N-PSDB; AAS65581.

XX New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits and to assess
 XX biodiversity

PS Claim 20: SEQ ID NO 31753; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
 XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
 XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 XX and gene mapping, and in recombinant production of (II). The
 XX polynucleotides are also used in diagnostics as expressed sequence tags
 XX for identifying expressed genes. (I) is useful in gene therapy techniques
 XX to restore normal activity of (II) or to treat disease states involving
 XX (II). (II) is useful for generating antibodies against it, detecting or
 XX quantitating a polypeptide in tissue, as molecular weight markers and as
 XX a food supplement. (II) and its binding partners are useful in medical
 XX imaging of sites expressing (II). (I) and (II) are useful for treating
 XX disorders involving aberrant protein expression or biological activity.
 XX The polypeptide and polynucleotide sequences have applications in
 XX diagnostics, forensics, gene mapping, identification of mutations
 XX responsible for genetic disorders or other traits to assess biodiversity
 XX and to produce other types of data and products dependent on DNA and
 XX amino acid sequences. ABG00010-ABG30377 represent novel human
 XX diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1062 AA;

Query Match

77.5%; Score 31; DB 22; Length 1062;

RESULT 35	
AAM64837	
ID	AAM64837 standard; Protein; 23 AA.
XX	
AC	AAM64837;

OS Homo sapiens.
XX
PN WO200157272-A2.
XX
PD 09-AUG-2001.

XX 30-JAN-2001; 2001WO-US00663.
XX 04-FEB-2000; 2000US-0180312.
XX 26-MAY-2000; 2000US-0207456.
XX 30-JUN-2000; 2000US-0608408.
XX 03-AUG-2000; 2000US-0632366.
XX 21-SEP-2000; 2000US-0234687.
XX 27-SEP-2000; 2000US-0236359.
XX 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488897/53.
XX Human genome-derived single exon nucleic acid probes useful for
XX analyzing gene expression in human placenta -
XX Claim 27; SEQ ID NO 38040; 654pp; English.
XX The present invention relates to single exon nucleic acid probes (SENP:
XX see AA131315-AA157346). The present sequence is a peptide encoded by one
XX such probe. The probes are useful for producing a microarray for
XX predicting, measuring and displaying gene expression in samples derived
XX from human placenta. The probes are useful for antenatal diagnosis of
XX human genetic disorders.
XX SQ Sequence 23 AA;
Query Match 75.0%; Score 30; DB 22; Length 23;
Best Local Similarity 85.7%; Pred. No. 37;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 3 LLPLQL 9
Db 11 LLPLQLL 17
RESULT 37
AAU01749
ID AAU01749-standard; Peptide; 37 AA.
XX AC AAU01749;
XX DT 18-JUL-2001 (first entry)
XX DE Human secreted protein #28.
XX KW Human; secreted protein; immunogen; antibody; diagnosis;
XX KW rheumatoid arthritis; hyperproliferative disorder; neoplasm;
XX KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;
XX KW angiogenesis; Alzheimer's disease; bacterial infection; viral infection;
XX KW fungal infection; corneal infection; wound healing; cell culture;
XX KW epithelial cell proliferation; skin ageing; transplantation;
XX KW tissue regeneration; chemotaxis; food additive.
XX OS Homo sapiens.
XX PN WO200123546-A1.
XX PD 05-APR-2001.
XX PF 26-SEP-2000; 2000WO-US26323.
XX PR 27-SEP-1999; 99US-0155805.
XX (HUMA-) HUMAN GENOME SCI INC.
XX PI Rosen CA, Ruben S, Komatsoulis GA;
XX WPI; 2001-266150/27.

DR N-PSDB; AAS02422.
XX Nucleic acids encoding 37 human secreted polypeptides, useful for
XX preventing, diagnosing and/or treating e.g. cancers, Parkinson's
XX disease and diabetic retinopathy -
XX Disclosure; Page 450; 494pp; English.
XX The sequence represents a human secreted protein of the invention. The
XX polynucleotides, polypeptides and antibodies raised against them are used
XX to prevent, treat or ameliorate a medical condition in e.g. humans, mice,
XX rabbits, goats, horses, cats, dogs, chickens or sheep. The
XX polynucleotides and antibodies are also used in diagnosing a pathological
XX condition or susceptibility to a pathological condition. The antibodies
XX can also be used in alleviating symptoms associated with the disorders
XX and in diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
XX immunosorbent assays (ELISA). Disorders which are diagnosed or treated
XX include autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX angiogenesis, nervous system disorders e.g. Alzheimer's disease, disorders
XX infections caused by bacteria, viruses and fungi and ocular disorders
XX e.g. corneal infection. The polypeptides can also be used to aid wound
XX healing and epithelial cell proliferation, to prevent skin aging due to
XX sunburn, to maintain organs before transplantation, for supporting cell
XX culture of primary tissues, to regenerate tissues and in chemotaxis. The
XX polypeptides can also be used as a food additive or preservative to
XX increase or decrease storage capabilities.
XX SQ Sequence 37 AA;
Query Match 75.0%; Score 30; DB 22; Length 37;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LLPLQL 9
Db 12 LLPLQLL 19
RESULT 38
ABB38810
ID ABB38810-standard; Peptide; 40 AA.
XX AC ABB38810;
XX DT 04-FEB-2002 (first entry)
XX DE Peptide #6316 encoded by human foetal liver single exon probe.
XX KW Human; foetal liver; gene expression; single exon nucleic acid probe.
XX OS Homo sapiens.
XX PN WO200157277-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00669.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-483447/52.

xx Human genome-derived single exon nucleic acid probes useful for
PT analyzing gene expression in human fetal liver -
xx Claim 27; SEQ ID NO 31445; 639pp + sequence listing; English.
xx The invention relates to a single exon nucleic acid probe for
CC measuring human gene expression in a sample derived from human foetal
CC liver. The single exon nucleic acid probes may be used for predicting,
CC measuring and displaying gene expression in samples derived from human
CC fetal liver. The present sequence is a peptide encoded by a single exon
CC nucleic acid probe of the invention.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
xx
SQ Sequence 40 AA;
Query Match 75.0%; Score 30; DB 22; Length 40;
Best Local Similarity 87.5%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LLLPLQIL 9
Db 29 LLLPLPIL 36
RESULT 39
ABB23836
ID ABB23836 standard; Protein; 40 AA.
xx
AC ABB23836;
xx
DT 23-JAN-2002 (first entry)
xx
DE Protein #5835 encoded by probe for measuring heart cell gene expression.
xx
KW Human; gene expression; heart; microarray; vascular system;
KW cardiovascular disease; hypertension; cardiac arrhythmia;
KW congenital heart disease.
xx
OS Homo sapiens.
xx
XX WO200157274-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00666.
PF
XX
XX 04-FEB-2000; 2000US-0180312.
PR
XX 26-MAY-2000; 2000US-0207456.
PR
XX 30-JUN-2000; 2000US-0608408.
PR
XX 03-AUG-2000; 2000US-0632366.
PR
XX 21-SEP-2000; 2000US-0234687.
PR
XX 27-SEP-2000; 2000US-0236359.
PR
XX 04-OCT-2000; 2000GB-0024263.
PR
xx (MOLE-) MOLECULAR DYNAMICS INC.
xx
xx Penn SG, Hanzel DK, Chen W, Rank DR;
xx
xx WPI; 2001-488899/53.
xx
xx Single exon nucleic acid probes for analyzing gene expression in human
xx hearts -
xx
xx Claim 15; SEQ ID No 25606; 530pp; English.
xx
xx The present invention relates to single exon nucleic acid probes for
xx measuring human gene expression in a sample derived from human heart (see
xx ABA21535-ABA41305). The present sequence is a protein encoded by one such
xx probe. The probes may be used for predicting, measuring and displaying
xx gene expression in samples derived from the human heart via microarrays.

CC By measuring gene expression, the probes are useful for predicting,
CC diagnosing, grading, staging, monitoring and prognosing diseases of the
CC human heart and vascular system e.g. cardiovascular disease,
CC hypertension, cardiac arrhythmias and congenital heart disease.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
xx
SQ Sequence 40 AA;
Query Match 75.0%; Score 30; DB 22; Length 40;
Best Local Similarity 87.5%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LLLPLQIL 9
Db 29 LLLPLPIL 36
RESULT 40
AAM59450
ID AAM59450 standard; Protein; 40 AA.
xx
AC AAM59450;
xx
XX 05-NOV-2001 (first entry)
DT
xx
DE Human brain expressed single exon probe encoded protein SEQ ID NO: 31555.
xx
KW Human; brain expressed exon; gene expression analysis; probe;
KW microarray; Alzheimer's disease; multiple sclerosis; schizophrenia;
KW epilepsy; cancer.
xx
OS Homo sapiens.
xx
XX WO200157275-A2.
PN
XX
XX 09-AUG-2001.
PD
XX
XX 30-JAN-2001; 2001WO-US00667.
PF
xx
xx 04-FEB-2000; 2000US-0180312.
PR
xx 26-MAY-2000; 2000US-0207456.
PR
xx 30-JUN-2000; 2000US-0608408.
PR
xx 03-AUG-2000; 2000US-0632366.
PR
xx 21-SEP-2000; 2000US-0234687.
PR
xx 27-SEP-2000; 2000US-0236359.
PR
xx 04-OCT-2000; 2000GB-0024263.
PR
xx (MOLE-) MOLECULAR DYNAMICS INC.
xx
xx Penn SG, Hanzel DK, Chen W, Rank DR;
xx
xx WPI; 2001-483446/52.
xx
xx Single exon nucleic acid probes for analyzing gene expression in human
xx brains -
xx
xx Example 4; SEQ ID NO: 31555; 650pp + Sequence Listing; English.
xx
xx The present invention provides a number of single exon nucleic acid
xx probes which are derived from genomic sequences expressed in the human
xx brain. They can be used to measure gene expression in brain cell samples,
xx which may enable the diagnosis and improved treatment of nervous system
xx diseases such as Alzheimer's disease, multiple sclerosis, schizophrenia,
xx epilepsy and cancers. The present sequence is a protein encoded by one of
xx the probes of the invention.
xx
SQ Sequence 40 AA;
Query Match 75.0%; Score 30; DB 22; Length 40;
Best Local Similarity 87.5%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy 2 LLLPLQL 9
   |||||
Db 29 LLLPLPIL 36

RESULT 41
AAM72012
ID AAM72012 standard; Protein; 40 AA.
XX AC AAM72012;
XX DT 06-NOV-2001 (first entry)
XX DE Human bone marrow expressed probe encoded protein SEQ ID NO: 32318.
XX KW Human: bone marrow expressed exon; gene expression analysis; probe;
XX KW microarray; cancer; leukaemia; lymphoma; myeloma.
XX OS Homo sapiens.
XX PN WO200157276-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00668.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488900/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human bone marrow -
XX PS Example 4; SEQ ID NO: 32318; 658pp + Sequence Listing; English.
XX CC The present invention provides a number of single exon nucleic acid
XX CC probes which are derived from genomic sequences expressed in the human
XX CC bone marrow. They can be used to measure gene expression in bone marrow
XX CC samples, which may enable the improved diagnosis and treatment of cancers
XX CC such as lymphoma, leukaemia and myeloma. The present sequence is a
XX CC protein encoded by one of the probes of the invention.
XX SQ Sequence 40 AA;
Query Match 75.0%; Score 30; DB 22; Length 40;
Best Local Similarity 87.5%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LLLPLQL 9
   |||||
Db 29 LLLPLPIL 36

RESULT 42
AAM19406
ID AAM19406 standard; Protein; 40 AA.
XX AC AAM19406;
XX DT 12-OCT-2001 (first entry)
XX DE Peptide #5840 encoded by probe for measuring cervical gene expression.
```

```
XX Probe; human; microarray; gene expression; cervical epithelial cell;
KW cervical cancer.
XX Homo sapiens.
XX PN WO200157278-A2.
XX PD 09-AUG-2001.
XX PF 30-JAN-2001; 2001WO-US00670.
XX PR 04-FEB-2000; 2000US-0180312.
XX PR 26-MAY-2000; 2000US-0207456.
XX PR 30-JUN-2000; 2000US-0608408.
XX PR 03-AUG-2000; 2000US-0632366.
XX PR 21-SEP-2000; 2000US-0234687.
XX PR 27-SEP-2000; 2000US-0236359.
XX PR 04-OCT-2000; 2000GB-0024263.
XX PA (MOLE-) MOLECULAR DYNAMICS INC.
XX PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX WPI; 2001-488901/53.
XX DR Human genome-derived single exon nucleic acid probes useful for
XX PT analyzing gene expression in human cervical epithelial cells -
XX PS Claim 27; SEQ ID NO 24232; 487pp; English.
XX CC The present invention relates to human single exon nucleic acid probes
XX CC (SENPs: see AAI10068-AAI28459). The present sequence is a peptide encoded
XX CC by one such probe. The SENPs are derived from human HeLa cells. The SENPs
XX CC can be used to produce a single exon microarray, which can be used for
XX CC measuring human gene expression in a sample derived from human cervical
XX CC epithelial cells. By measuring gene expression, the probes are therefore
XX CC useful in grading and/or staging of diseases of the cervix, notably
XX CC cervical cancer.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX SQ Sequence 40 AA;
Query Match 75.0%; Score 30; DB 22; Length 40;
Best Local Similarity 87.5%; Pred. No. 65;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 2 LLLPLQL 9
   |||||
Db 29 LLLPLPIL 36

RESULT 43
AAM32284
ID AAM32284 standard; Protein; 40 AA.
XX AC AAM32284;
XX DT 17-OCT-2001 (first entry)
XX DE Peptide #6321 encoded by probe for measuring placental gene expression.
XX KW Probe; microarray; human; placenta; antenatal diagnosis;
XX KW genetic disorder.
XX OS Homo sapiens.
XX PN WO200157272-A2.
XX PD 09-AUG-2001.
XX
```

[illegible]

```
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155148.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 12-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
```

```
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 75.0%; Score 30; DB 21; Length 70;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 SLLPLQI 8
Db 18 ALLPLEI 25
:||||:|

RESULT 45
AAM84905
ID AAM84905 standard; Protein; 83 AA.
XX
AC AAM84905;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human immune/haematopoietic antigen SEQ ID NO:12498.
XX
KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis.
XX
OS Homo sapiens.
XX
PN WO200157182-A2.
XX
PD 09-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01354.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 11-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
```

PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-483426/52.
DR N-PSDB; AAK57686.
XX
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
XX metastasis -
XX
XX Claim 11; SEQ ID NO 12498; 3071pp + Sequence Listing; English.
XX
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patients own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169

CC represent sequences used in the exemplification of the present invention.

XX

SQ Sequence 83 AA;

Query Match 75.0%; Score 30; DB 22; Length 83;

Best Local Similarity 85.7%; Pred. No. 1.4e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 LLPLQIL 9

|||||

Db 36 LLPLQLL 42

Search completed: November 6, 2002, 12:05:11

Job time : 24.7778 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 18.4444 Seconds
(without alignments)
84.413 Million cell updates/sec

Title: US-09-905-083-80
Perfect score: 50
Sequence: 1 GPLVCRGTL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	45	90.0	251	11 Q9DBQ8	Q9dbq8 mus musculus
2	41	82.0	87	11 Q9CQ78	Q9cq78 mus musculus
3	41	82.0	249	11 Q9QYN4	Q9qyn4 mus musculus
4	41	82.0	276	11 Q9QYN3	Q9qyn3 mus musculus
5	41	82.0	3201	2 Q9F828	Q9f828 micromonospora
6	41	82.0	3546	2 Q9F830	Q9f830 micromonospora
7	40	80.0	254	6 Q9XSN6	Q9xsn6 sus scrofa
8	40	80.0	262	2 Q9XAC9	Q9xac9 streptomyces
9	39	78.0	155	17 Q9YBC2	Q9ybc2 streptomyces
10	39	78.0	247	13 Q9W7Q5	Q9w7q5 paralichthy
11	39	78.0	288	5 Q9W174	Q9w174 drosophila
12	38	76.0	119	4 Q9NR68	Q9nr68 homo sapien
13	38	76.0	234	11 Q9CV76	Q9cv76 mus musculus
14	38	76.0	234	11 Q9R048	Q9r048 mus musculus
15	38	76.0	249	11 Q91VE3	Q91ve3 mus musculus
16	38	76.0	254	11 Q61096	Q61096 mus musculus

17	38	76.0	514	11 Q9CS05	Q9cs05 mus musculus
18	37	74.0	46	11 Q62540	Q62540 mus spretus
19	37	74.0	108	11 Q9CVU2	Q9cvu2 mus musculus
20	37	74.0	114	11 Q9D974	Q9d974 mus musculus
21	37	74.0	117	13 Q9PUF3	Q9puf3 bothrops ja
22	37	74.0	128	11 Q9D6K1	Q9d6k1 mus musculus
23	37	74.0	135	11 Q62284	Q62284 mus musculus
24	37	74.0	204	4 Q96JE1	Q96je1 homo sapien
25	37	74.0	205	4 Q96JE2	Q96je2 homo sapien
26	37	74.0	205	4 Q96JE0	Q96je0 homo sapien
27	37	74.0	219	13 Q91036	Q91036 gadus morhua
28	37	74.0	235	11 Q63274	Q63274 rattus norv
29	37	74.0	239	11 Q63275	Q63275 rattus norv
30	37	74.0	246	11 Q88301	Q88301 mus musculus
31	37	74.0	247	5 Q18458	Q18458 heterodera
32	37	74.0	250	11 Q03955	Q03955 praomys nat
33	37	74.0	253	11 Q91Y82	Q91y82 mus musculus
34	37	74.0	255	4 Q96RQ0	Q96rq0 homo sapien
35	37	74.0	256	11 Q9R0K0	Q9r0k0 mus musculus
36	37	74.0	261	6 Q29474	Q29474 canis famli
37	37	74.0	261	11 Q9JM70	Q9jm70 mus musculus
38	37	74.0	263	11 Q9JM71	Q9jm71 mus musculus
39	37	74.0	311	2 Q9AQ55	Q9ags5 burkholderi
40	37	74.0	599	10 Q9STY4	Q9sty4 arabidopsis
41	36	72.0	66	11 Q61752	Q61752 mus musculus
42	36	72.0	138	11 Q9D725	Q9d725 mus musculus
43	36	72.0	149	11 Q925V8	Q925v8 mus musculus
44	36	72.0	178	13 Q93594	Q93594 dicentrarch
45	36	72.0	233	13 Q9PT51	Q9pt51 agkistrodon

ALIGNMENTS

RESULT 1

Q9DBQ8	PRELIMINARY;	PRT;	251 AA.
ID	Q9DBQ8		
AC	Q9DBQ8;		
DT	01-JUN-2001 (TREMBLrel. 17, Created)		
DT	01-JUN-2001 (TREMBLrel. 17, Last sequence update)		
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)		
DE	1200016C12RIK PROTEIN.		
GN	1200016C12RIK.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=C57BL/6J; TISSUE=LUNG;		
RX	MEDLINE=21085660; PubMed=11217851;		
RA	Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,		
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,		
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,		
RA	Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R.,		
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,		
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,		
RA	Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,		
RA	Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,		
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,		
RA	Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,		
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,		
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamlya M., Lee N.H.,		
RA	Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,		
RA	Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,		
RA	Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,		
RA	Suzuki H., Tovo-Oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,		
RA	Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,		
RA	Hayashizaki Y.;		
RT	*Functional annotation of a full-length mouse cDNA collection.*;		
RL	Nature 409:685-690(2001).		
CC	!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE		
CC	TRYPSIN FAMILY.		

DR EMBL; AK004807; BAB23579.1; -
 DR HSSP; P00763; 1DPO.
 DR MGD; MGI:1921082; 1200016C12Rik.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 251 AA; 28153 MW; F4D667F8C80C4A23 CRC64;

Query Match 90.0%; Score 45; DB 11; Length 251;
 Best Local Similarity 88.9%; Pred. No. 0.49;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Caps 0;

QY 1 GPLVCRGTL 9
 DB 206 GPLVCEGTL 214
 ||||| |||

RESULT 2
 Q9CQ78 PRELIMINARY; PRT; 87 AA.
 AC Q9CQ78;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE 2310015108RIK PROTEIN (FRAGMENT).
 GN PRSS20 OR 2310015108RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Brownstein M.J., Bult C., Fletcher C., Carninci P., de Bonaldo M.F.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Wynshaw-Boris A., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; AB016226; BAA88825.1; -
 DR EMBL; AK009720; BAB26461.1; -
 DR EMBL; AK009360; BAB26241.1; -
 DR HSSP; P00763; 1DPO.
 DR MGD; MGI:1929977; Prss20.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 87 AA; 9549 MW; B988D0CD62926EAA CRC64;

Query Match 82.0%; Score 41; DB 11; Length 87;

Best Local Similarity 77.8%; Pred. No. 1.1;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 GPLVCRGTL 9
 DB 42 GPLVCGSL 50
 ||||| |

RESULT 3
 Q9QYN4 PRELIMINARY; PRT; 249 AA.
 AC Q9QYN4;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE HIPPOSTASIN (2310015108RIK PROTEIN).
 GN PRSS20 OR 2310015108RIK.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]

SEQUENCE FROM N.A.
 RC TISSUE=BRIN;
 RA Yamaguchi N., Mitsui S.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]

SEQUENCE FROM N.A.
 RC TISSUE=BRIN;
 RA Mitsui S., Yamaguchi N.;
 RL "cDNA cloning of a novel brain serine protease, Hippostasin.";
 RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]

SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=TONGUE;
 RX MEDLINE=21085660; PubMed=11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
 RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
 RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
 RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
 RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
 RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
 RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
 RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
 RA Brownstein M.J., Bult C., Fletcher C., Carninci P., de Bonaldo M.F.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Wynshaw-Boris A., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
 RA Hayashizaki Y.;
 RT Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL; AB016226; BAA88825.1; -
 DR EMBL; AK009720; BAB26461.1; -
 DR EMBL; AK009360; BAB26241.1; -
 DR HSSP; P00763; 1DPO.
 DR MGD; MGI:1929977; Prss20.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease.


```
FT      Protein      /label= Pro-peptide
FT      8..244
FT      /label= Mat_protein
XX
XX      WQ9530758-A1.
XX
XX      16-NOV-1995.
XX
XX      09-MAY-1995; 95WO-0506157.
XX
XX      02-MAY-1995; 95US-0427767.
XX      10-MAY-1994; 94US-0241174.
XX
XX      (HYBR-) HYBRITECH INC.
XX      (MAYO-) MAYO FOUNDATION.
XX
XX      Saedi MS, Tindall DJ, Young CYF;
XX
XX      WPI; 1995-404123/51.
XX      N-PSDB; AAT05149.
XX
XX      New isolated prostate-specific kallikrein polypeptide(s) - used to
XX      develop prods. for use in assays for such polypeptide(s), partic.
XX      for diagnosis and monitoring of prostate cancer
XX
XX      Disclosure; Page 39-40; 61pp; English.
XX
XX      Human pro-hk2 (AAR84669) is a precursor form of a prostate-specific
XX      glandular kallikrein (AAR84667) useful in the development of prods. used
XX      in the diagnosis and monitoring of prostate cancer. Pro-hk2 is obtd.
XX      by expression in prokaryotic (E. coli) or eukaryotic (insect, mammalian)
XX      cells of cDNA (AAT05149) derived from human benign prostatic hyperplasia
XX      tissue.
XX
XX      Sequence 244 AA;
XX
XX      Query Match 80.0%; Score 40; DB 16; Length 244;
XX      Best Local Similarity 77.8%; Pred. No. 38;
XX      Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX      QY 1 GPLVCRGTL 9
XX      | | | | | | |
XX      Db 198 GPLVCNGVL 206
XX
XX      RESULT 42
XX      AAW83204
XX      ID AAW83204 standard; Protein; 244 AA.
XX      AC AAW83204;
XX      DT 09-FEB-1999 (first entry)
XX      DE Prostate-specific glandular kallikrein protein phk2.
XX      KW Prostate cancer; detection; hk2; hk1; hk3; phk2; tissue kallikrein;
XX      KWK2; prostate-specific glandular kallikrein protein; PSA; human;
XX      KW prostate-specific antigen.
XX      OS Homo sapiens.
XX      PN WQ9846795-A1.
XX      PR 22-OCT-1998.
XX      XX
XX      PF 09-APR-1998; 98WO-0507027.
XX      PR 11-APR-1997; 97US-0843076.
XX      XX
XX      PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX      (MAYO-) MAYO FOUNDATION.
XX
XX      PI Slawin KM, Tindall DJ, Young CYF;
```

```
XX
XX      WPI; 1998-594592/50.
XX      N-PSDB; AAV70341.
XX
XX      Detection of human kallikrein 2 RNA - by reverse transcription and
XX      amplification by PCR, for detecting, monitoring and staging of
XX      prostate cancer
XX
XX      Example 1; Fig 2; 90pp; English.
XX
XX      The present invention describes a diagnostic method for detecting human
XX      kallikrein 2 (hk2) DNA. The method comprises: (a) contacting DNA obtained
XX      by reverse transcription (RT) of RNA from a human physiological sample
XX      which comprises cells suspected of containing hk2 RNA with at least 2
XX      oligonucleotides to amplify the DNA by PCR to yield amplified hk2 DNA,
XX      where the conditions amplify the DNA obtained by RT of RNA from at least
XX      one cell containing hk2 in a sample which comprises at least 107 to 109
XX      cells; and (b) detecting the presence of the amplified hk2 DNA. The
XX      method can be used for detecting, monitoring the progression of and
XX      pathologically staging prostate cancer. The present sequence represents
XX      prostate-specific glandular kallikrein protein (phk2).
XX
XX      Sequence 244 AA;
XX
XX      Query Match 80.0%; Score 40; DB 19; Length 244;
XX      Best Local Similarity 77.8%; Pred. No. 38;
XX      Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX      QY 1 GPLVCRGTL 9
XX      | | | | | | |
XX      Db 198 GPLVCNGVL 206
XX
XX      RESULT 43
XX      AAW45396
XX      ID AAW45396 standard; Protein; 244 AA.
XX      AC AAW45396;
XX      DT 06-JUL-1998 (first entry)
XX      DE Prostate-specific glandular kallikrein precursor pro-hk2.
XX      KW Prostate-specific glandular kallikrein; phk2 protein; hk2; antigen;
XX      KW prostate carcinoma; prostate cancer; benign prostatic hyperplasia;
XX      KW diagnosis; human.
XX      OS Homo sapiens.
XX      FH Key
XX      FT Peptide 1..7
XX      FT Protein 8..244
XX      FT /label= Mat_protein
XX      PN WQ9802748-A1.
XX      XX
XX      PD 22-JAN-1998.
XX      XX
XX      PF 15-JUL-1997; 97WO-US12322.
XX      PR 15-JUL-1996; 96US-0680868.
XX      XX
XX      PA (HYBR-) HYBRITECH INC.
XX      (MAYO-) MAYO FOUNDATION.
XX      XX
XX      PI Grauer L, Klee GG, Mikolajczyk SD, Saedi M, Tindall DJ;
XX      PI Young CYF;
XX      XX
XX      WPI; 1998-120378/11.
XX      N-PSDB; AAV06603.
XX
XX      Diagnostic methods using antibodies which bind prostate antigens -
```

PT useful for, e.g. monitoring treatment or progression of prostate
 XX cancer
 PS
 XX Example 1; Page 51-52; 100pp; English.
 XX
 CC This polypeptide comprises human prostate-specific glandular
 CC kallikrein precursor pro-hk2 (phk2). It is encoded by cDNA (see
 CC AAV06603) derived from human benign prostate hyperplasia (BPH) tissue
 CC RNA. The invention provides a diagnostic method comprising
 CC contacting antibodies that specifically bind to pro-hk2 or mature
 CC hk2 (see AAV45395) with a sample of physiological fluid from a human.
 CC The assay is based on the discovery that phk2 is detected in the
 CC supernatant of a prostate carcinoma cell line and that hk2 is
 CC present in human physiological fluid from prostate cancer cells.
 CC The method is useful for monitoring the treatment and/or
 CC progression of prostate cancer, or for the early detection of
 CC prostate cancer in males that have BPH or a high grade prostatic
 CC neoplasia (HPGN) or whose family members have or had BPH, HPGN or
 CC a prostate cancer. Monitoring the presence and/or amount of hk2
 CC complexes with plasma proteins may also be important in
 CC distinguishing between prostate cancer and BPH.
 XX
 XX Sequence 244 AA;

Query Match 80.0%; Score 40; DB 19; Length 244;
 Best Local Similarity 77.8%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 | | | | | | |
 Db 198 GPLVCGVL 206

RESULT 44
 AAW96188
 ID AAW96188 standard; Protein; 244 AA.

AC AAW96188;

DT 27-APR-1999 (first entry)

DE Pro human Kallikrein 2 (prohk2).

XX Human kallikrein 2; hk2; breast cancer; diagnosis; detection;
 KW treatment; monitoring; prostate specific antigen; PSA.
 XX

OS Homo sapiens.

XX WO9859073-A1.

XX 30-DEC-1998.

XX 19-JUN-1998; 98WO-US12840.

XX 20-JUN-1997; 97US-0050963.

XX (KLEE/) KLEE G G.

PA (NAYO-) MAYO FOUNDATION.

PA (TIND/) TINDALL D J.

PA (YOUN/) YOUNG C Y F.

XX Klee GG, Tindall DJ, Young C Y F;

XX WPI; 1999-105632/09.

DR N-PSDB; AAX08946.

XX Use of human kallikrein 2 - as a marker for developing products for
 PT the diagnosis, prognosis, monitoring and treatment of breast cancer
 XX

PS Disclosure; Page 35-36; 67pp; English.

XX Human kallikrein 2 (hk2) is expressed at elevated levels relative to
 CC the prostate cancer antigen, prostate specific antigen (PSA) by

CC breast cancer cells when stimulated by androgens. Detecting levels
 CC of hk2 may provide a simple diagnostic tool for detecting or
 CC determining breast cancer. Detecting hk2 is achieved by producing
 CC an hk2 DNA from hk2 RNA by reverse transcription. The hk2 DNA
 CC obtained is then amplified by PCR and detected using probes.
 CC Determination of breast cancer may also be determined by exposing
 CC the hk2 polypeptide to an agent which binds to it and then detecting
 CC the binary complex formed. The amount of complex formed correlates
 CC directly to the presence of breast cancer in the human individual.
 CC The methods can be used in the diagnosis, treatment and/or
 CC monitoring of the progression or remission of breast cancer and/or
 CC monitoring hk2 levels.
 XX

SQ Sequence 244 AA;

Query Match 80.0%; Score 40; DB 20; Length 244;

Best Local Similarity 77.8%; Pred. No. 38;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 | | | | | | |

Db 198 GPLVCGVL 206

RESULT 45

AAW03130

ID AAW03130 standard; Protein; 250 AA.

XX AC AAW03130;

XX DT 21-OCT-1996 (first entry)

XX DE Prostate-specific antigen.

XX Prostate-specific antigen; metastasis; carcinoma; tumour; cancer;

KW neoplasm; archival tissue; polymerase chain reaction; RT-PCR;

KW primer; kallikrein.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..13

FT /label= Pro_peptide

FT Protein 14..250

FT /label= Mat_protein

FT /note= "prostate-specific antigen"

XX WO9621042-A2.

XX 11-JUL-1996.

XX 04-JAN-1996; 96WO-US00461.

XX 04-JAN-1995; 95US-0368706.

XX (UYBO-) UNIV BOSTON.

XX Edelman RA, Moreland RB;

XX WPI; 1996-334017/33.

DR N-PSDB; AAT35867.

XX Detecting a metastatic disorder in a fixed biological sample -
 PT comprises reverse transcription PCR of sample nucleic acid and
 PT detection of metastatic, esp. prostate, tissue-specific target
 PT sequence

XX Disclosure; Fig 1; 60pp; English.

XX Prostate-specific antigen (PSA) (AAW03130) is a serine protease
 CC (kallikrein) found in both normal and neoplastic prostate
 CC epithelium. It can be used as a prostate marker. PCR primers
 CC (see also AAT35855-62) based on PSA cDNA (AAT35867) or genomic

CC (AAT35868) sequences can used for the specific amplification of PSA
CC nucleic acids in fresh and archival tissue samples as a means of
CC detecting metastatic sequences.

XX
SQ Sequence 250 AA;

Query Match 80.0%; Score 40; DB 17; Length 250;
Best Local Similarity 77.8%; Pred. No. 39;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
| | | | | | |
Db 204 GPLVCGVL 212

Search completed: November 6, 2002, 12:05:13
Job time : 25.7778 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 18.4444 Seconds
(without alignments)
84.413 Million cell updates/sec

Title: us-09-905-083-86
Perfect score: 41
Sequence: 1 MARSLLLPL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues
Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_virus.*
- 16: sp_bacteriopl.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	85.4	396	4	Q9NZL1
2	35	85.4	468	4	Q9NZL2
3	35	85.4	1192	10	Q9S7T0
4	33	80.5	272	10	Q9FNC4
5	33	80.5	397	16	Q9AA49
6	33	80.5	526	10	Q9ZRH9
7	33	80.5	571	10	Q9SYC9
8	33	80.5	934	13	Q9DER4
9	33	80.5	2998	5	Q9NKS4
10	32	78.0	81	16	Q9ZG73
11	32	78.0	126	16	Q9KRE6
12	32	78.0	227	4	Q9UKJ0
13	32	78.0	247	2	Q9KWM3
14	32	78.0	271	4	Q9HBS0
15	32	78.0	303	4	Q9UKJ1
16	32	78.0	306	6	Q95KJ4

17	32	78.0	334	16	Q988I7
18	32	78.0	334	16	Q9HVB1
19	32	78.0	506	16	Q9Z5I3
20	32	78.0	654	6	Q95LX1
21	32	78.0	654	6	Q95LV1
22	32	78.0	921	10	Q9SDQ4
23	32	78.0	923	10	Q9ZV68
24	31	75.6	138	11	Q60463
25	31	75.6	147	2	Q44395
26	31	75.6	223	4	Q96AD7
27	31	75.6	227	8	Q9XPF7
28	31	75.6	237	4	Q9BSQ8
29	31	75.6	294	10	Q9LQR9
30	31	75.6	308	16	Q9ZRW2
31	31	75.6	375	16	Q25341
32	31	75.6	375	16	Q9ZLL1
33	31	75.6	439	2	P96097
34	31	75.6	452	16	Q9KF24
35	31	75.6	481	5	Q9NII3
36	31	75.6	708	2	O87239
37	31	75.6	5069	2	O52545
38	31	75.6	5069	2	O52789
39	30	73.2	74	12	O84649
40	30	73.2	140	2	O9KGS7
41	30	73.2	146	11	O35972
42	30	73.2	146	11	O63750
43	30	73.2	161	8	O9T9G7
44	30	73.2	170	10	O38778
45	30	73.2	207	8	O21512

ALIGNMENTS

RESULT 1
Q9NZL1 PRELIMINARY: PRT; 396 AA.
ID Q9NZL1
AC Q9NZL1
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ASPARTYL PROTEASE.
DE BACE2.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20422477; PubMed=10965118;
RA Solans A., Estivill X., de La Luna S.;
RT "A new aspartyl protease on 21q22.3, BACE2, is highly similar to
RT Alzheimer's amyloid precursor protein beta-secretase."
RL Cytochrome. Cell Genet. 89:177-184(2000).
DR EMBL; AF188277; AAF35836.1; -
DR HSSP; P00797; 2REN.
DR MEROPS; A01.041; -
DR InterPro; IPR001969; Asp_protease.
DR InterPro; IPR001481; Pepsin.
DR Pfam; PF000026; asp; 2.
DR PRINTS; PR00792; PEPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 396 AA; 43013 MW; 5023A7AF39ICEAC9 CRC64;

Query Match 85.4%; Score 35; DB 4; Length 396;
Best Local Similarity 77.8%; Pred. No. 25;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
:||:|||||
Db 4 LARALLPL 12

```
RESULT 2
Q9NZL2
ID Q9NZL2 PRELIMINARY; PRT; 468 AA.
AC
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE ASPARTYL PROTEASE.
GN BACE2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=20422477; PubMed=10965118;
RA Solans A., Estivill X., de la Luna S.;
RT "A new aspartyl protease on 21q22.3, BACE2, is highly similar to
RT Alzheimer's amyloid precursor protein beta-secretase.";
RL CytoGenet. Cell Genet. 89:177-184(2000).
DR EMBL; AF188276; AAF35835.1; -
DR HSP; P00797; 2REN.
DR MEROPS; A01.04; -.
DR InterPro; IPR001969; Asp_protease.
DR Pfam; PF00026; asp; 3.
DR PRINTS; PR00792; PPSIN.
DR PROSITE; PS00141; ASP_PROTEASE; UNKNOWN_2.
KW Protease.
SQ SEQUENCE 468 AA; 50324 MW; 717E0920126A0142 CRC64;

Query Match 85.4%; Score 35; DB 4; Length 468;
Best Local Similarity 77.8%; Pred. No. 29;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
:|||||
Db 4 LARALLPL 12

RESULT 3
Q9S7T0 PRELIMINARY; PRT; 1192 AA.
AC Q9S7T0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE F28J7.11 PROTEIN.
GN F28J7.11 OR F4P13.33.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Roning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F28J7 genomic sequence.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Roning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome III BAC F4P13 genomic sequence.";
RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC010797; AAF03433.1; -
DR EMBL; AC009325; AAF01560.1; -
```

```
DR InterPro; IPR002048; EF-hand.
DR PROSITE; PS00018; EF_HAND; UNKNOWN 1.
SQ SEQUENCE 1192 AA; 132863 MW; 4F67B124CBAFF154 CRC64;

Query Match 85.4%; Score 35; DB 10; Length 1192;
Best Local Similarity 77.8%; Pred. No. 76;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
|||||
Db 237 MARSLVLPV 245

RESULT 4
Q9FNC4 PRELIMINARY; PRT; 272 AA.
ID Q9FNC4
AC Q9FNC4;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE VEGETATIVE STORAGE PROTEIN-LIKE (PUTATIVE VEGETATIVE STORAGE
DE PROTEIN).
GN AT5G44020 OR MRH10.13.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98069011; PubMed=9405937;
RA Kotani H., Nakamura Y., Sato S., Kaneko T., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. II.
RT Sequence features of the regions of 1,044,062 bp covered by thirteen
RT physically assigned p1 clones.";
RL DNA Res. 4:291-300(1997).
RN (2)
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
RN (3)
RP SEQUENCE FROM N.A.
RA Yamada K., Liu S.X., Pham P.K., Banh J., Dale J.M., Goldsmith A.D.,
RA Jiang P.X., Lee J.M., Onodera C.S., Quach H.L., Tang C., Toriumi M.,
RA Yamamura Y., Yu G., Yu S., Bowser L., Carninci P., Chen H., Cheuk R.,
RA Hayashizaki Y., Ishida J., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Kim C., Koesema E., Lam B., Lin J., Meyers M.C., Miranda M.,
RA Narusaka M., Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M.,
RA Shinn P., Southwick A., Tracy S.E., Shinozaki K., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Full length cDNA of gene MRH10.13/AT5G44020 (GI:9758561).";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB006703; BAB09062.1; -
DR EMBL; AY054542; AAK96733.1; -
DR EMBL; AY045786; AAK76460.1; -
SQ SEQUENCE 272 AA; 31056 MW; ESFB016FDAFOEA93 CRC64;
```

```
Query Match 80.5%; Score 33; DB 10; Length 272;
Best Local Similarity 88.9%; Pred. No. 45;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 MARSLLLPL 9
|||||
Db 1 MARSLLLSL 9
```

RESULT 5

Q9AA49 ID Q9AA49 PRELIMINARY; PRT; 397 AA.
AC Q9AA49;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN CC0757.
GN CC0757.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; alpha subdivision; Caulobacter group;
OC Caulobacter.
OX NCBI_TaxID=69394;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647;
RA Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A., Stephens C., Phadke N.D., Ely B.,
RA DeBoy R.T., Dodson R.J., Durkin A.S., Gwinn M.L., Haft D.H.,
RA Kolonay J.F., Smit J., Craven M.B., Khouri H., Shetty J., Berry K.,
RA Utterback T., Tran K., Wolf A., Vamathevan J., Ermolaeva M., White O.,
RA Salzberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AB005751; AAK22742.1; -;
DR TIGR; CC0757; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 397 AA; 42687 MW; AF0EB915C02ED58B CRC64;

Query Match 80.5%; Score 33; DB 16; Length 397;
Best Local Similarity 66.7%; Pred. No. 66;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
:|:|:|:|
Db 285 LARALILPL 293

RESULT 6

Q9ZRH9 ID Q9ZRH9 PRELIMINARY; PRT; 526 AA.
AC Q9ZRH9;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE EARLY EMBRYOGENESIS PROTEIN.
GN OSE351.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. TAINUNG 67;
RA Tseung M.J., Wang C.S., Hsu H.R.;
RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; U25968; AAD10369.1; -;
DR InterPro; IPR000459; Seedstore.lls.
DR Pfam; PF00190; Seedstore.lls; 4.
DR PRINTS; PR00439; 1LSGLOBULIN.
SQ SEQUENCE 526 AA; 58770 MW; BC23C0F9D1711F9B CRC64;

Query Match 80.5%; Score 33; DB 10; Length 526;
Best Local Similarity 77.8%; Pred. No. 88;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
|||:|:|:|
Db 6 MAASLLIPL 14

RESULT 7

Q9SYC9 ID Q9SYC9 PRELIMINARY; PRT; 571 AA.
AC Q9SYC9;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F1M15.13 PROTEIN.
GN F1M15.13
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC euroids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Altafi H., Araujo R., Huizar L., Rowley D., Buehler E., Dunn P.,
RA Gonzalez A., Kremetskaia I., Kim C., Lenz C., Li J., Liu S.,
RA Luros S., Schwartz J., Shinn P., Toriumi M., Vysotskaia V.S.,
RA Walker M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC006085; AAD30639.1; -;
DR InterPro; IPR000535; MSP_domain.
DR InterPro; IPR00157; TIR.
DR Pfam; PF00635; MSP_domain; 2.
DR Pfam; PF01582; TIR; 1.
DR SMART; SM00255; TIR; 1.
SQ SEQUENCE 571 AA; 64407 MW; 61B6C9F20696FFE7 CRC64;

Query Match 80.5%; Score 33; DB 10; Length 571;
Best Local Similarity 77.8%; Pred. No. 96;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
|:|:|:|:|
Db 332 MLRSLLVPL 340

RESULT 8

Q9DER4 ID Q9DER4 PRELIMINARY; PRT; 934 AA.
AC Q9DER4;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ZONA PELLUCIDA PROTEIN 1.
GN ZP1.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20435849; PubMed=10979984;
RA Bausek N., Wacławek M., Schneider W.J., Wohlrab F.;
RT "The major chicken egg envelope protein ZP1 is different from ZPB and
RT is synthesized in the liver.";
RL J. Biol. Chem. 275:28866-28872(2000).
DR EMBL; AJ289697; CAC16087.1; -;
DR HSP; P04155; LPS2.
DR InterPro; IPR000519; P_trefoil.
DR InterPro; IPR001507; zona_pellucida.
DR Pfam; PF00100; zona_pellucida; 1.
DR PRINTS; PR00023; ZPELLUCIDA.
DR SMART; SM00018; P; 1.
DR SMART; SM00241; ZP; 1.
SQ SEQUENCE 934 AA; 100028 MW; C28CB7743245B47F CRC64;

Query Match 80.5%; Score 33; DB 13; Length 934;
Best Local Similarity 87.5%; Pred. No. 1.6e+02;

Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
:|||||
Db 4 SRSLLLPL 11

RESULT 9

Q9NKS4 PRELIMINARY; PRT; 2998 AA.
AC Q9NKS4;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE L505.2.
GN L505.2.
OS Leishmania major.
OC Eukaryota; Euzoenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIDLIN;
RA Myler P.J.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC005892; AAF35915.1; -.
DR InterPro: IPR000560; His_acid_phosphatase.
DR PROSITE: PS00778; HIS_ACID_PHOSPHAT_2; UNKNOWN1.
SQ SEQUENCE 2998 AA; 311341 MW; B303361F7E0B1786 CRC64;

Query Match 80.5%; Score 33; DB 5; Length 2998;
Best Local Similarity 87.5%; Pred. No. 5.2e+02;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
:|||||
Db 37 ARSVLLPL 44

RESULT 10

Q92G73 PRELIMINARY; PRT; 81 AA.
AC Q92G73;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN RC1250.
GN RC1250.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OX Rickettsiaceae; Rickettsieae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MALISH 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissbach J., Claverie J.-M.,
RA Raoult D.;
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii.";
RL Science 293:2093-2098 (2001).
DR EMBL; A6008672; AAL03788.1; -.
KW Hypothetical protein; Complete proteome.

Query Match 78.0%; Score 32; DB 16; Length 81;
Best Local Similarity 75.0%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
:|||||
Db 43 ARSLLIPI 50

RESULT 11

Q9KRE6 PRELIMINARY; PRT; 126 AA.
AC Q9KRE6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE DNA-BINDING PROTEIN INHIBITOR ID-2-RELATED PROTEIN.
GN VC1696.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
cholerae.";
RL Nature 406:477-483 (2000).
DR EMBL; AE004247; AAF94846.1; -.
DR TIGR; VC1696; -.

Query Match 78.0%; Score 32; DB 16; Length 126;
Best Local Similarity 87.5%; Pred. No. 34;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 8
:|||||
Db 33 MLRSLLPL 40

RESULT 12

Q9UKJ0 PRELIMINARY; PRT; 227 AA.
AC Q9UKJ0;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ACTIVATING RECEPTOR PILRBETA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20127940; PubMed=10660620;
RA Mousseau D.D., Banville D., L'Abbe D., Bouchard P., Shen S.H.;
RT "PILRalpha, a novel immunoreceptor tyrosine-based inhibitory motif-
bearing protein, recruits SHP-1 upon tyrosine phosphorylation and is
paired with the truncated counterpart PILRBeta.";
RL J. Biol. Chem. 275:4467-4474 (2000).
DR EMBL; AF161081; AAD52965.1; -.
DR InterPro: IPR003599; Ig.
DR SMART; SM00409; IG; 1.
KW Receptor.

Query Match 78.0%; Score 32; DB 4; Length 227;
Best Local Similarity 77.8%; Pred. No. 61;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
:|||||
Db 1 MGRPLLLPL 9


```

RESULT 13
Q9KWM3 PRELIMINARY; PRT; 247 AA.
AC Q9KWM3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NUKM, NUKA, ORF1, ORF3, ORF4, ORF5, ORF6 GENES, PARTIAL AND COMPLETE
DE CDS.
OS Staphylococcus warneri.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1292;
RN [1]
RP SEQUENCE FROM N.A.
RA Sashihara T., Kimura H., Higuchi T., Matsusaki H., Sonomoto K.,
RA Ishizaki A.;
RT "Staphylococcus warneri truncated nukM, nukA, orf1 genes, and the
RT downstream region.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Kimura H., Matsusaki H., Sashihara T., Sonomoto K., Ishizaki A.;
RT "Purification and partial identification of bacteriocin ISK-1, a new
RT lantibiotic produced by pediococcus sp. ISK-1.";
RL Biosci. Biotechnol. Biochem. 62:2341-2345(1998).
CC -|- SIMILARITY: THE N-TERMINAL REGION IS SIMILAR TO THAT OF OTHER
CC REGULATORY COMPONENTS OF SENSORY TRANSDUCTION SYSTEMS.
DR EMBL; AB034941; BAA95675.1; -.
DR InterPro: IPR001789; Response_reg.
DR Pfam: PF000072; response_reg; 1.
DR SMART: SM00448; REC; 1.
DR Phosphorylation; Sensory transduction.
KW SEQUENCE 247 AA; 28805 MW; E5E5070475479089 CRC64;

Query Match 78.0%; Score 32; DB 2; Length 247;
Best Local Similarity 66.7%; Pred. No. 67;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 193 VSRSLLIPL 201

RESULT 14
Q9HBS0 PRELIMINARY; PRT; 271 AA.
AC Q9HBS0;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 29.6 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,
RA Yu J., Han L.H.;
RT "Novel human cDNA clones with function of inhibiting cancer cell
RT growth.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF217981; AAG17224.1; -.
KW Hypothetical protein.
SQ SEQUENCE 271 AA; 29630 MW; 31158B0F8F03B41F CRC64;

Query Match 78.0%; Score 32; DB 4; Length 271;
Best Local Similarity 77.8%; Pred. No. 73;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 MARSLLLPL 9
Db 123 MGRPLLLPL 131

RESULT 15
Q9UKJ1 PRELIMINARY; PRT; 303 AA.
AC Q9UKJ1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE INHIBITORY RECEPTOR PILRALPHA.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20127940; PubMed=10660620;
RA Mousseau D.D., Banville D., L'Abbe D., Bouchard P., Shen S.H.;
RT "PILRALpha, a novel immunoreceptor tyrosine-based inhibitory motif-
RT bearing protein, recruits SHP-1 upon tyrosine phosphorylation and is
RT paired with the truncated counterpart PILRbeta.";
RL J. Biol. Chem. 275:4467-4474(2000).
DR EMBL; AF161080; AAD52964.1; -.
DR InterPro: IPR003599; Ig.
DR SMART; SM00409; IG; 1.
KW Receptor.
SQ SEQUENCE 303 AA; 33877 MW; 0410ADF7E80928B CRC64;

Query Match 78.0%; Score 32; DB 4; Length 303;
Best Local Similarity 77.8%; Pred. No. 82;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 1 MGRPLLLPL 9

RESULT 16
Q95KJ4 PRELIMINARY; PRT; 306 AA.
AC Q95KJ4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 35.7 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TEMPORAL LOBE RIGHT;
RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,
RA Suzuki Y., Sugano S., Hashimoto K.;
RT "Isolation of full-length cDNA clones from macaque brain cDNA
RT libraries.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB060830; BAB46862.1; -.
KW Hypothetical protein.
SQ SEQUENCE 306 AA; 35721 MW; CEE59B99CE0DC4F CRC64;

Query Match 78.0%; Score 32; DB 6; Length 306;
Best Local Similarity 66.7%; Pred. No. 83;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 254 LARGILLPL 262

```

```

Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ARSLLLPL 9
Db 245 ARELLPL 252

RESULT 19
Q925I3 PRELIMINARY; PRT; 506 AA.
ID Q925I3
AC Q925I3;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 53.9 KDA PROTEIN (CONSERVED MEMBRANE PROTEIN).
GN ML1544 OR MLCB596.27.
OS Mycobacterium leprae.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]
RP SEQUENCE FROM N.A.
RA Oliver K., Harris D.;
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=21128732; PubMed=11234002;
RA Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajadream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus.";
RL Nature 409:1007-1011(2001).
DR EMBL; AL035472; CAB36583.1; -
DR EMBL; AL583922; CAC30495.1; -
DR Leproma; ML1544; -
DR KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 506 AA; 53881 MW; D8D329345C27EAD4 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 506;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
Db 223 MNRSVLLPL 231

RESULT 20
Q95LY1 PRELIMINARY; PRT; 654 AA.
ID Q95LY1
AC Q95LY1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 74.1 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;

Best Local Similarity 87.5%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ARSLLLPL 9
Db 18 ARSILMPL 25

RESULT 18
Q9HVB1 PRELIMINARY; PRT; 432 AA.
ID Q9HVB1
AC Q9HVB1;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HYPOTHETICAL PROTEIN PA4684.
GN PA4684.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Ruffnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004882; AAG08071.1; -
RW Hypothetical protein; Complete proteome.
SQ SEQUENCE 432 AA; 49194 MW; 3010520E9BA694B3 CRC64;

Query Match 78.0%; Score 32; DB 16; Length 432;
```

RT "Isolation of novel full-length cDNA clones from macaque testis cdna
libraries."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB071060; BAB64453.1; -
KW Hypothetical protein
SQ SEQUENCE 654 AA; 74081 MW; 148CE97DCD1F8AA4 CRC64;

Query Match 78.0%; Score 32; DB 6; Length 654;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLPL 9
|||||
DB 10 RSLLPL 16

RESULT 21
Q95LV1 PRELIMINARY; PRT; 654 AA.
AC Q95LV1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PUTATIVE LIGAND-GATED ION CHANNEL SUBUNIT.
OS HYPOTHETICAL 74.1 KDA PROTEIN.
OC Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cdna
libraries."
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB071090; BAB64484.1; -
KW Hypothetical protein.
SQ SEQUENCE 654 AA; 74112 MW; E0815FB08856919E CRC64;

Query Match 78.0%; Score 32; DB 6; Length 654;
Best Local Similarity 100.0%; Pred. No. 1.8e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLPL 9
|||||
DB 10 RSLLPL 16

RESULT 22
Q9SD04 PRELIMINARY; PRT; 921 AA.
AC Q9SD04;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DR GLR5.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA; TISSUE=SEEDLING;
RA Davenport R.J., Kiegle E.A., Tester M.;
RT "GLR5, an ionotropic glutamate receptor ortholog from Arabidopsis."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF210701; AAF21042.1; -
DR InterPro: IPR001828; ANF_receptor.
DR InterPro: IPR001320; Ion_glut_receptor.
DR InterPro: IPR001311; SBP_glut_receptor.

DR Pfam: PF01094; ANF_receptor; 2.
DR Pfam: PF00060; lig_chan; 1.
DR SMART: SM00079; PBPe; 1.
SQ SEQUENCE 921 AA; 103491 MW; 69F9707A4D63C55D CRC64;

Query Match 78.0%; Score 32; DB 10; Length 921;
Best Local Similarity 77.8%; Pred. No. 2.5e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
|||||
DB 705 MARSRLVPL 713

RESULT 23
Q9ZV68 PRELIMINARY; PRT; 923 AA.
AC Q9ZV68;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE LIGAND-GATED ION CHANNEL SUBUNIT.
GN AT2G32400.
OS Arabidopsis thaliana (Mouse-ear cross).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Bueli C.R., Ketchum K.A., Lee J.J., Ronning C.M., Roo H., Moffat K.S.,
RA Cronin L.A., Shen M., VanAken S.E., Umayam L., Tallon L.J., Gill J.E.,
RA Adams M.D., Carrera A.J., Creasy T.H., Goodman H.M., Somerville C.R.,
RA Copenhaver G.P., Preuss D., Nierman W.C., White O., Eisen J.A.,
RA Salzberg S.L., Fraser C.M., Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana."
RT Nature 402:761-768(1999).
RL Nature 402:761-768(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC005700; AAC69938.1; -
DR InterPro: IPR001828; ANF_receptor.
DR InterPro: IPR001320; Ion_glut_receptor.
DR InterPro: IPR001311; SBP_glut_receptor.
DR Pfam: PF01094; ANF_receptor; 2.
DR Pfam: PF00060; lig_chan; 1.
DR SMART: SM00079; PBPe; 1.
SQ SEQUENCE 923 AA; 103671 MW; EA79BC9AB96838B6 CRC64;

Query Match 78.0%; Score 32; DB 10; Length 923;
Best Local Similarity 77.8%; Pred. No. 2.6e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
|||||
DB 707 MARSRLVPL 715

RESULT 24
Q60463 PRELIMINARY; PRT; 138 AA.
AC Q60463;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SODIUM CHANNEL ALPHA SUBUNIT (FRAGMENT).

GN CHOL.
OS Cricetus longicaudatus (Long-tailed hamster) (Chinese hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Cricetus.
OX NCBI_TaxID=10030;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=OVARY;
RX MEDLINE=93235989; PubMed=7682773;
RA Lalik P.H., Krafft D.S., Ciccarelli R.B.;
RT "Characterization of endogenous sodium channel gene expressed in
Chinese hamster ovary cells."
RL An. J. Physiol. 264:C803-C809(1993).
DR EMBL; M87540; AAA36978.1; -
KW Ionic channel.
FT NON_TER 138 138
SQ SEQUENCE 138 AA; 15638 MW; 6CC9B20CE472A470 CRC64;

Query Match 75.6%; Score 31; DB 11; Length 138;
Best Local Similarity 75.0%; Pred. No. 60;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLP 8
Db [1] [1] [1] [1]
1 MARSVLVP 8

RESULT 25
Q44395 PRELIMINARY; PRT; 147 AA.
AC Q44395;
DT 01-NOV-1996 (TRENBLrel. 01, Created)
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE TI PLASMID PT15955 T-DNA REGION.
OS Agrobacterium tumefaciens.
OG Plasmid Ti.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=358;
RN [1]
RP SEQUENCE FROM N.A.
RA Barker R.F., Idler K.B., Thompson D.V., Kemp J.D.;
RT "Nucleotide sequence of the T-DNA region from the Agrobacterium
tumefaciens octopine Ti plasmid pTi15955."
RL Plant Mol. Biol. 2:335-350(1983).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93225814; PubMed=8469115;
RA Turk S.C.H.J., Nester E.W., Hooykaas P.J.J.;
RT "The virA promoter is a host-range determinant in Agrobacterium
tumefaciens."
RL Mol. Microbiol. 7:719-724(1993).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=94035196; PubMed=8220492;
RA Guevara-Garcia A., Mosqueda-cano G., Arguuello-Astorga G., Simpson J.,
RA Herrera-Estrella L.;
RT "Tissue-specific and wound-inducible pattern of expression of the
magnopine synthase promoter is determined by the interaction between
positive and negative cis-regulatory elements."
RL Plant J. 4:495-505(1993).
DR EMBL; X00493; CAA25178.1; -
KW Plasmid.
SQ SEQUENCE 147 AA; 16844 MW; 0A4651F0C613F95 CRC64;

Query Match 75.6%; Score 31; DB 2; Length 147;
Best Local Similarity 87.5%; Pred. No. 64;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLP 8
Db [1] [1] [1] [1]

Db 1 MARYLLLP 8

RESULT 26
Q96AD7 PRELIMINARY; PRT; 223 AA.
AC Q96AD7;
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE HYPOTHETICAL 24.2 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRIN, AND NEUROBLASTOMA;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017264; AAH17264.1; -
KW Hypothetical protein.
SQ SEQUENCE 223 AA; 24152 MW; 3178AA3E46191803 CRC64;

Query Match 75.6%; Score 31; DB 4; Length 223;
Best Local Similarity 87.5%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLP 9
Db [1] [1] [1] [1]
66 ARPLLPL 73

RESULT 27
Q9APF7 PRELIMINARY; PRT; 227 AA.
AC Q9APF7;
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
DE ATP SYNTHASE A CHAIN (EC 3.6.1.34).
GN ATPASE6.
OS Gonostoma gracile.
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Stomiliformes; Gonostomatidae; Gonostoma.
OX NCBI_TaxID=48457;
RN [1]
RP SEQUENCE FROM N.A.
RX PubMed=10525676;
RA Miya M., Nishida M.;
RT "Organization of the mitochondrial genome of a deep-sea fish,
Gonostoma gracile (Teleostei: Stomiliformes): First example of transfer
RNA gene rearrangements in bony fishes.";
RL Mar. Biotechnol. 1:416-426(1999).
CC -1- FUNCTION: KEY COMPONENT OF THE PROTON CHANNEL; IT MAY PLAY A
DIRECT ROLE IN THE TRANSLOCATION OF PROTONS ACROSS THE MEMBRANE
(BY SIMILARITY).
CC -1- SUBUNIT: F-TYPE ATPASES HAVE 2 COMPONENTS, CF(1) - THE CATALYTIC
CORE - AND CF(0) - THE MEMBRANE PROTON CHANNEL. CF(1) HAS FIVE
SUBUNITS: ALPHA(3), BETA(3), GAMMA(1), DELTA(1), EPSILON(1). CF(0)
HAS THREE MAIN SUBUNITS: A, B AND C (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ATPASE A CHAIN FAMILY.
DR EMBL; AB016274; BAA82490.1; -
DR HSSP; P00855; 1c17.
DR InterPro: IPR000568; ATP_synt_A.
DR Pfam: PF00119; ATP_synt_A; 1.
DR PRINTS; P00123; ATPASEA.
DR PROSITE; PS00449; ATPASEA; 1.
KW CF(0); Hydrogen ion transport; Mitochondrion; Transmembrane.
SQ SEQUENCE 227 AA; 24753 MW; 453EDBF00622BA8A CRC64;

```
Query Match 75.6%; Score 31; DB 8; Length 227;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 0; Gaps 0;

QY 1 MARSLLLPL 9
   :|: |||||
DB 54 LAKQLLLPL 62

RESULT 28
Q9BSQ8          PRELIMINARY;      PRT;    237 AA.
ID Q9BSQ8;
AC Q9BSQ8;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 25.5 KDA PROTEIN (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN, NEUROBLASTOMA;
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004895; AAH04895.1; -.
KW Hypothetical protein.
FT NON_TER
SQ SEQUENCE 237 AA; 25537 MW; 09DA9B5DA39846A1 CRC64;

Query Match 75.6%; Score 31; DB 4; Length 237;
Best Local Similarity 87.5%; Pred. No. 1e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
   || |||||
DB 80 ARPLLLPL 87

RESULT 29
Q9LOK9          PRELIMINARY;      PRT;    294 AA.
ID Q9LOK9;
AC Q9LOK9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F5D14.26 PROTEIN.
GN F5D14.26.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Liu S.-X., Chan A., Yu G., Lee J.-M., Lenz C., Pham P., Sakano H.,
RA Toriumi M., Vysotskaia V.S., Chin C., Chiou J., Choi E., Chung M.,
RA Gonzalez A., Howing B., Liu A., Vaysberg M., Altafi H., Brooks S.,
RA Buehler E., Chao Q., Conn L., Conway A.B., Hansen N.F.,
RA Johnson-Hopson C., Khan S., Kim C., Lam B., Miranda M., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Davis R.W., Ecker J.R.,
RA Federspiel N.A., Theologis A.;
RT "The sequence of BAC F5D14 from Arabidopsis thaliana chromosome 1.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007767; AAF81346.1; -.

DR HSP; P08200; IISO.
SQ SEQUENCE 294 AA; 32851 MW; 46575E963FAE9C68 CRC64;

Query Match 75.6%; Score 31; DB 10; Length 294;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
   | | |||||
DB 249 MLRHLLPL 257

RESULT 30
Q92R72          PRELIMINARY;      PRT;    308 AA.
ID Q92R72;
AC Q92R72;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN SMC02382.
GN SMC02382.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Ramsperger U., Surzycki R., Thebault P., Vandenbol M.,
RA Vorhoelter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL Science 293:668-672(2001).
DR EMBL; AL591785; CAC45610.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 308 AA; 33322 MW; 58B756C82EE20D89 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 308;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
   :||| | | |
DB 97 VARSLKLPL 105

RESULT 31
Q25341          PRELIMINARY;      PRT;    375 AA.
ID Q25341;
AC Q25341;
DT 01-JAN-1998 (TrEMBLrel. 05, Created)
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SOLUTE-BINDING SIGNATURE AND MITOCHONDRIAL SIGNATURE PROTEIN (ASPB).
GN HP0624.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=210;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=26695 / ATCC 700392;
RX MEDLINE=97394467; PubMed=9252185;
RA Tomb J.-F., White O., Kervage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
```

```

RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khalak H.G., Glodek A.,
RA McKenney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.K.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RA "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori.";
RL Nature 388:539-547(1997).
DR EMBL: AE000576; AAD07687.1; -.
DR TIGR: HP0624; -.
DR InterPro: IPR001511; AminoTran_1.
DR InterPro: IPR000567; SBP_bac_1.
DR Pfam: PF00155; aminoTran_1_2; 1.
DR PROSITE; PS01037; SBP_BACTERIAL_1; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 375 AA; 43124 MW; 262EE78985E56683 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 375;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
Db 129 AKSLMLPL 136
1:|||||

RESULT 32
Q9ZLL1 PRELIMINARY; PRT; 375 AA.
AC Q9ZLL1;
DT 01-MAY-1999 (TrEMBLrel. 10, Created)
DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AMINOTRANSFERASE.
GN JHP0568.
OS Helicobacter pylori J99 (Campylobacter pylori J99).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
OX NCBI_TaxID=85963;
RN 1;
RP SEQUENCE FROM N.A.
RX MEDLINE=99120557; PubMed=9923682;
RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
RA Trust T.J.;
RT "Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen Helicobacter pylori.";
RL Nature 397:176-180(1999).
DR EMBL: AE001489; AAD06139.1; -.
DR InterPro: IPR001176; ACC_synthase.
DR InterPro: IPR001511; AminoTran_1.
DR InterPro: IPR000567; SBP_bac_1.
DR Pfam: PF00155; aminoTran_1_2; 1.
DR PRINTS; PR00753; ACCSYNTHASE.
DR PROSITE; PS01037; SBP_BACTERIAL_1; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 375 AA; 43036 MW; E39C31B43C6D3924 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 375;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
Db 129 AKSLMLPL 136
1:|||||

```

```

RESULT 33
P96097

```

```

ID P96097 PRELIMINARY; PRT; 439 AA.
AC P96097;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 50.8 KDA PROTEIN.
OS Thiobacillus ferrooxidans.
OG Plasmid pTF5.
OC Bacteria; Proteobacteria; gamma subdivision; Acidithiobacillus.
OX NCBI_TaxID=920;
RN 1;
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC33020;
RX MEDLINE=98015401; PubMed=9353917;
RA Dominy C.N., Deane S.M., Rawlings D.E.;
RT "A geographically widespread plasmid from Thiobacillus ferrooxidans
RT has genes for ferredoxin-, FNR-, prismsane- and NADH-oxidoreductase-
RT like proteins which are also located on the chromosome.";
RL Microbiology 143:0-0(0).
DR EMBL: U73041; AAC80177.1; -.
DR InterPro: IPR002559; Transposase_11.
DR Pfam: PF01609; Transposase_11; 1.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 439 AA; 50811 MW; ACCEC49B30B36551 CRC64;

Query Match 75.6%; Score 31; DB 2; Length 439;
Best Local Similarity 77.8%; Pred. No. 2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 193 MRRRLPL 201
1:|||||

RESULT 34
Q9KF24 PRELIMINARY; PRT; 452 AA.
ID Q9KF24;
AC Q9KF24;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PTS SYSTEM, N-ACETYLGLUCOSAMINE-SPECIFIC ENZYME II, ABC COMPONENT
DE (EIIABC-NAG).
GN BH0673.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN 1;
RP SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL: AP001509; BAB04392.1; -.
DR HSP: P05053; IIBA.
DR InterPro: IPR001996; PTS_EIIB.
DR InterPro: IPR003352; PTS_EIIC.
DR Pfam: PF00367; PTS_EIIB; 1.
DR Pfam: PF02378; PTS_EIIC; 1.
KW Complete proteome.
SQ SEQUENCE 452 AA; 48918 MW; 78E0248E868C0033 CRC64;

Query Match 75.6%; Score 31; DB 16; Length 452;
Best Local Similarity 66.7%; Pred. No. 2e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
1:|||||

```

```
Db 8 IGRSLLPLI 16

RESULT 35
Q9NLI3 PRELIMINARY; PRT; 481 AA.
AC Q9NII3
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HIGH-AFFINITY NA+-DEPENDENT GLUTAMATE TRANSPORTER.
GN EAAT1.
OS Diploptera punctata (Pacific beetle cockroach).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Orthopteroidea; Dictyoptera; Blattaria;
OC Blaberoides; Blaberidae; Diploptera.
OX NCBI_TaxID=6984;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RX MEDLINE=20211311; PubMed=10745160;
RA Donly C., Jevnikar J., McLean H., Caveney S.;
RT "Substrate-stereoselectivity of a high-affinity glutamate transporter
RT cloned from the CNS of the cockroach Diploptera punctata.";
RL Insect Biochem. Mol. Biol. 30:369-376(2000).
DR EMBL; AF208521; AAF71701.1; -.
DR InterPro; IPR001991; Na_dicarboxyl_symp.
DR Pfam; PF00375; SDR; 1.
DR PRINTS; PRO0173; EDTNSPORT.
DR PROSITE; PS00714; NA_DICARBOXYL_SYMP_2; 1.
DR SEQUENCE 481 AA; 52404 MW; C5B61EA9B2CD2E44 CRC64;
SQ SEQUENCE 481 AA; 52404 MW; C5B61EA9B2CD2E44 CRC64;

Query Match 75.6%; Score 31; DB 5; Length 481;
Best Local Similarity 66.7%; Pred.No. 2.1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLPL 9
Db 62 MKSLPL 70

RESULT 36
O87239 PRELIMINARY; PRT; 708 AA.
AC O87239
DT 01-NOV-1998 (TREMBLrel. 08, Created)
DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE CYTOLYSIN B TRANSPORT PROTEIN.
OS Lactococcus lactis.
OG Plasmid pmrc01.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DPC3147;
RX MEDLINE=99000510; PubMed=9767571;
RA Dougherty B.A., Hill C., Weidman J.F., Richardson D.R., Venter J.C.,
RA Ross R.P.;
RT "Sequence and analysis of the 60 kb conjugative, bacteriocin-producing
RT plasmid pmrc01 from Lactococcus lactis DPC3147.";
RL Mol. Microbiol. 29:1029-1038(1998).
DR EMBL; AE001272; AAC56012.1; -.
DR MEROPS; C39; UPW; -.
DR InterPro; IPR003593; AAA.
DR InterPro; IPR001140; ABC_transporter_tmern.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR001687; ATP_GTP_A.
DR Pfam; PF00664; ABC_membrane; 1.
DR Pfam; PF00005; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR ATP-binding; Plasmid.
KW
```

DR PROSITE; PS00606; B_KETOACYL-SYNTHASE; 2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
 KW Phosphopantetheine; Transferase.
 SQ SEQUENCE 5060 AA; 526542 MW; 13A6EBE1A48A6B0 CRC64;

Query Match 75.6%; Score 31; DB 2; Length 5060;
 Best Local Similarity 66.7%; Pred. No. 2.4e+03;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
 |||||:
 Db 4871 MAESLLVPI 4879

RESULT 38

ID 052789 PRELIMINARY; PRT; 5069 AA.
 AC 052789;
 DT 01-JUN-1998 (TrEMBLrel. 06, Created)
 DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE RIFAMYCIN POLYKETIDE SYNTHASE, TYPE 1
 OS Anycolatopsis mediterranei (Nocardia mediterranei).
 OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
 OC Actinomycetales; Pseudonocardineae; Pseudonocardaceae; Amycolatopsis.
 OX NCBI_TaxID=33910;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LBG A3136;
 RA Schupp T., Toupet C., Engel N., Goff S.;
 RT "Cloning and sequence analysis of the putative rifamycin polyketide
 RT synthase gene cluster from Amycolatopsis mediterranei.";
 RL Submitted (DEC-1997) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ223012; CAAL1036.1;
 DR HSSP: P25715; 1MLA.
 DR InterPro: IPR001227; Acyltransf_domain.
 DR InterPro: IPR003781; DUF184.
 DR InterPro: IPR000794; Ketoacyl-synt.
 DR InterPro: IPR003880; Phosphopant_attach.
 DR Pfam: PF00698; Acyl_transf; 3.
 DR Pfam: PF02629; GOA_binding; 1.
 DR Pfam: PF00109; ketoacyl-synt; 3.
 DR Pfam: PF02801; ketoacyl-synt_C; 3.
 DR Pfam: PF00550; pp-binding; 3.
 DR PROSITE; PS00075; ACP_DOMAIN; 3.
 DR PROSITE; PS00606; B_KETOACYL-SYNTHASE; 2.
 DR PROSITE; PS00012; PHOSPHOPANTETHEINE; 3.
 KW Phosphopantetheine; Transferase.
 SQ SEQUENCE 5069 AA; 527694 MW; C6E545CB77C0AAMF CRC64;

Query Match 75.6%; Score 31; DB 2; Length 5069;
 Best Local Similarity 66.7%; Pred. No. 2.4e+03;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
 |||||:
 Db 4880 MAESLLVPI 4888

RESULT 39

ID 084649 PRELIMINARY; PRT; 74 AA.
 AC 084649;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE A335L PROTEIN.
 GN A335L.
 OS Paramoecium bursaria chlorella virus 1 (PBCV-1).
 OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
 OX NCBI_TaxID=10506;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE-96187795; PubMed-8614977;
 RA Lu Z., Li Y., Que Q., Kutish G.F., Rock D.L., Van Etten J.L.;
 RT "Analysis of 94 kb of the chlorella virus PBCV-1 330-kb genome: map
 RT positions 88 to 182";
 RL Virology 216:102-123(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20013326; PubMed-10544099;
 RA Kaiser A., Vollmert M., Tholl D., Graves M.V., Gurnon J.R., Xing W.,
 RA Liseac A.D., Nickerson K.W., Van Etten J.L.;
 RT "Chlorella virus PBCV-1 encodes a functional homospermidine
 RT synthase.";
 RL Virology 263:254-262(1999).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-20478054; PubMed-11021991;
 RA Sun L., Gurnon J.R., Adams B.J., Graves M.V., Van Etten J.L.;
 RT "Characterization of a beta-1,3-glucanase encoded by chlorella virus
 RT PBCV-1";
 RL Virology 276:27-36(2000).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Van Etten J.L.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Van Etten J.L.;
 RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Van Etten J.L.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP SEQUENCE FROM N.A.
 RA Graves M.V., Van Etten J.L.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [8]
 RP SEQUENCE FROM N.A.
 RA Graves M.V., Van Etten J.L.;
 RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [9]
 RP SEQUENCE FROM N.A.
 RA Gurnon J.R., Graves M.V., Van Etten J.L.;
 RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: U42580; AAC96703.1;
 SQ SEQUENCE 74 AA; 8674 MW; E374DA3D333A73ED CRC64;

Query Match 73.2%; Score 30; DB 12; Length 74;
 Best Local Similarity 66.7%; Pred. No. 52;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
 :|:||||
 Db 19 LPRNLLPL 27

RESULT 40

ID 09KGS7 PRELIMINARY; PRT; 140 AA.
 AC 09KGS7;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE APRI.
 GN APRI.
 OS Pseudomonas brassicaecarum.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=86264;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NFM421;
 RX MEDLINE-21125577; PubMed-11222613;

RA Chabeaud P., de Groot A., Bitter W., Tommassen J., Heulin T.,
 RA Achouak W.;
 RT "Phase-Variable Expression of an Operon Encoding Extracellular
 RT Alkaline Protease, a Serine Protease Homolog, and Lipase in
 RT Pseudomonas brassicaecarum.";
 RL J. Bacteriol. 183:2117-2120(2001).
 DR EMBL; AF286062; AAF87589.1; -;
 DR HSSP; P18958; 1SMP.
 SQ SEQUENCE 140 AA; 15267 MW; 95E0B2761C7E03C9 CRC64;

Query Match 73.2%; Score 30; DB 2; Length 140;
 Best Local Similarity 100.0%; Pred. No. 1e+02;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLL 7
 Db 24 MARSLLL 30

RESULT 41
 O35972 PRELIMINARY; PRT; 146 AA.
 AC O35972;
 DT 01-JAN-1998 (TREMBLrel. 05, Created)
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE RIBOSOMAL PROTEIN L23 (L23 MITOCHONDRIAL-RELATED PROTEIN) (L23MRP)
 DE (MITOCHONDRIAL RIBOSOMAL PROTEIN L23) (L23MT).
 GN MRPL23 OR RPL23L OR RPL23 OR L23MRP OR MRPL23.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=I29SVJ; TISSUE=LIVER;
 RX MEDLINE=98008916; PubMed=9344651;
 RA Zubair M., Hilton K., Saam J.R., Surani M.A., Tilghman S.M.,
 RA Sasaki H.;
 RT "Structure and expression of the mouse L23mrp gene downstream of the
 RT imprinted H19 gene: biallelic expression and lack of interaction with
 RT the H19 enhancers.";
 RL Genomics 45:290-296(1997).
 RN [2]
 RP SEQUENCE OF 8-44 FROM N.A.
 RC STRAIN=C57BL/6;
 RA Greally J.M., Guinness M., McGrath J., Zemel S.;
 RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Suzuki T., Terasaki M., Takemoto C., Hanada T., Wada A., Ueda T.,
 RA Watanabe K.;
 RT "Mammalian mitochondrial ribosome; structural and functional
 RT compensation for deficit of RNA with protein components.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U84902; AAC53392.1; -;
 DR EMBL; U84903; AAC53393.1; -;
 DR EMBL; U71209; AAB70459.1; -;
 DR EMBL; AB049646; BAB40851.1; -;
 DR MGD; MGI:1196612; Mrpl23.
 DR InterPro; IPR001014; Ribosomal_L23.
 DR Pfam; PF00276; Ribosomal_L23; 1.
 KW Ribosomal protein.
 SQ SEQUENCE 146 AA; 17122 MW; 0116ADE56C9B1553 CRC64;

Query Match 73.2%; Score 30; DB 11; Length 146;
 Best Local Similarity 66.7%; Pred. No. 1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
 Db 1 MARNVLYPL 9

RESULT 42
 O63750 PRELIMINARY; PRT; 146 AA.
 ID O63750
 AC O63750;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE RIBOSOMAL PROTEIN L23-RELATED PRODUCT HOMOLOG.
 GN RL23MRP.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Qian N.;
 RL Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL; U62635; AAB05795.1; -;
 DR InterPro; IPR001014; Ribosomal_L23.
 DR Pfam; PF00276; Ribosomal_L23; 1.
 SQ SEQUENCE 146 AA; 17050 MW; 1A0CICEE119D7534 CRC64;

Query Match 73.2%; Score 30; DB 11; Length 146;
 Best Local Similarity 66.7%; Pred. No. 1e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
 Db 1 MARNVLYPL 9

RESULT 43
 Q9T9G7 PRELIMINARY; PRT; 161 AA.
 ID Q9T9G7
 AC Q9T9G7;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE ND6 PROTEIN.
 GN ND6.
 OS Pupa strigosa.
 OC Mitochondrion.
 OC Eukaryota; Metazoa; Mollusca; Gastropoda; Opisthobranchia;
 OC Cephalaspidea; Acteonacea; Acteonidae; Pupa.
 OX NCBI_TaxID=98460;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20142339; PubMed=10677849;
 RA Kurabayashi A., Ueshima R.;
 RT "Complete sequence of the mitochondrial DNA of the primitive
 RT opisthobranch gastropod Pupa strigosa: systematic implication of the
 RT genome organization.";
 RL Mol. Biol. Evol. 17:266-277(2000).
 DR EMBL; AB028237; BAA89016.1; -;
 DR InterPro; IPR001457; Oxidored_q3.
 DR Pfam; PF00499; oxidored_q3; 1.
 KW Mitochondrion.
 SQ SEQUENCE 161 AA; 17332 MW; F732ECA81C25382C CRC64;

Query Match 73.2%; Score 30; DB 8; Length 161;
 Best Local Similarity 77.8%; Pred. No. 1.2e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
 Db 13 LAISLLLPL 21

RESULT 44
 Q38778 PRELIMINARY; PRT; 170 AA.
 ID Q38778
 AC Q38778;

DT 01-NOV-1996 (TReMBLrel. 01, Created)
DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE DNA-BINDING PROTEIN.
GN PF-1.
OS Avena sativa (Oat).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae;
OC Poaceae; Avena.
OX NCBI_TaxID=4498;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GARY;
RX MEDLINE=94203798; PubMed=8152915;
RA Nieto-Sotelo J., Ichida A., Quail P.H.;
RT "Positive Factor 1 (PF-1) from oat is an HMGV- and H1 histone-like
RT protein that binds a functionally defined AT-rich DNA element in the
RT oat phtochrome A gene (PHYA3) promoter.";
RL Nucleic Acids Res. 22:1115-1116(1994).
DR EMBL; L24391; AAA32718.1; -;
DR InterPro: IPR000637; AT_hook.
DR InterPro: IPR001386; Linkerhist_N.
DR Pfam; PF02178; AT_hook; 3.
DR Pfam; PF00538; linker_histone; 1.
DR PRINTS; PR00929; ATHOOK.
DR ProDom; PD000373; Linkerhist_N; 1.
DR SMART; SM00384; AT_hook; 3.
DR SMART; SM00526; H15; 1.
KW DNA-binding.
SQ SEQUENCE 170 AA; 18057 MW; CB26DAB1DE4347D5 CRC64;

Query Match 73.2%; Score 30; DB 10; Length 170;

Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLP 8
DB 6 VAKSLLLP 13

RESULT 45

O21512
ID O21512 PRELIMINARY; PRT; 207 AA.
AC O21512;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE NADH-UBIQUINONE OXIDOREDUCTASE CHAIN 4 (EC 1.6.5.3) (FRAGMENT).
GN ND4.
OS Acomys cahirinus (Egyptian spiny mouse).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Acomys.
OX NCBI_TaxID=10068;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98152303; PubMed=9491603;
RA Engel S.R., Hogan K.M., Taylor J.F., Davis S.K.;
RT "Molecular systematics and paleobiogeography of the South American
RT sigmodontine rodents.";
RL Mol. Biol. Evol. 15:35-49(1998).
CC -1- CATALYTIC ACTIVITY: NADH + UBIQUINONE = NAD(+) + UBIQUINOL.
DR EMBL; U83803; AAB87156.1; -;
DR InterPro: IPR001750; Oxidored_q1.
DR InterPro: IPR000260; Oxidored_q5_N.
DR Pfam; PF00361; oxidored_q1; 1.
DR Pfam; PF01059; oxidored_q5_N; 1.
KW Mitochondrion; NAD; Oxidoreductase; Ubiquinone.
FT NON_TER 207
SQ SEQUENCE 207 AA; 23915 MW; 01C11B6E1A142B8F CRC64;

Query Match 73.2%; Score 30; DB 8; Length 207;

Best Local Similarity 66.7%; Pred. No. 1.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MARSLLLP 9
DB 6 MASTMLLP 14
Search completed: November 6, 2002, 12:12:15
Job time : 22.4444 secs

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 5.33333 Seconds
(without alignments)
65.339 Million cell updates/sec

Title: US-09-905-083-86
Perfect score: 41
Sequence: 1 MARSLILPL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	253	1 KLK7_HUMAN	P49862 homo sapien
2	35	85.4	518	1 BAE2_HUMAN	Q9Y320 homo sapien
3	34	82.9	461	1 SSL1_YEAST	Q04673 saccharomyc
4	33	80.5	464	1 OTSA_RHISN	P55612 rhizobium s
5	33	80.5	582	1 MM14_MOUSE	P53690 mus musculu
6	33	80.5	582	1 MM14_RAT	Q10739 rattus norv
7	32	78.0	580	1 MM14_PIG	Q9XT90 sus scrofa
8	32	78.0	1025	1 SNXD_HUMAN	Q9Y3W8 homo sapien
9	32	78.0	1628	1 YABE_SCHPO	Q09779 schizosacch
10	31	75.6	169	1 SULA_SALTY	P08847 salmonella
11	31	75.6	346	1 NU2M_EQUAS	P92476 equus asinu
12	31	75.6	2005	1 CIN2_RAT	P04775 rattus norv
13	30	73.2	72	1 COR2_DROME	Q26377 drosophila
14	30	73.2	148	1 CYTC_RABIT	Q97862 oryctolagus
15	30	73.2	299	1 YHJC_ECOLI	P37641 escherichia
16	30	73.2	354	1 CD68_HUMAN	P34810 homo sapien
17	30	73.2	357	1 CYHV_GIBBA	P32963 gibberella
18	30	73.2	575	1 PTL1_ECOLI	P08839 escherichia
19	30	73.2	575	1 PTL1_HAEIN	P43922 haemophilus
20	30	73.2	575	1 PTL1_SALTY	P12654 salmonella
21	30	73.2	1318	1 VIVD_BPT7	P03726 bacterioph
22	30	73.2	2554	1 7LES_DROME	P13368 drosophila
23	29	70.7	129	1 IGF2_MUSVI	P41694 mustela vis
24	29	70.7	167	1 QCRA_BACSU	P46911 bacillus su
25	29	70.7	260	1 NIRO_PSEAE	Q51481 pseudomonas
26	29	70.7	276	1 PYGI_SYNEL	P50039 synechococc
27	29	70.7	328	1 TH12_SCHPO	P40998 schizosacch
28	29	70.7	370	1 CLCB_PSEPU	P11452 pseudomonas
29	29	70.7	463	1 HSLU_THEMEA	Q9WY22 thermotoga
30	29	70.7	555	1 INUL_KLUMA	P28999 kluyveromyc
31	29	70.7	705	1 MMLD_STRCO	Q9XAB6 streptomyce
32	29	70.7	731	1 CATA_HALMA	O59651 haloarcula
33	29	70.7	992	1 PHS2_DICDI	P34114 dictyosteli

RESULT 1				
ID	KLK7_HUMAN	STANDARD;	PRT;	253 AA.
AC	P49862;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic enzyme) (hSCCE).			
DE	KLK7 OR PRSS6 OR SCCE.			
GN	Homo sapiens (Human).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.			
RC	TISSUE=Skin;			
RC	MEDLINE=94308225; PubMed=8034709;			
RX	Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A.,			
RA	Egelrud T.;			
RA	"Cloning, expression, and characterization of stratum corneum			
RT	chymotryptic enzyme. A skin-specific human serine proteinase.";			
RT	J. Biol. Chem. 269:19420-19426(1994).			
RL	[2]			
RN	SEQUENCE FROM N.A.			
RP	TISSUE=Keratinocytes;			
RC	Yousef G.M., Scorillas A., Diamandis E.P.;			
RA	"Molecular characterization, mapping and tissue expression of the			
RT	human stratum corneum chymotryptic enzyme gene.";			
RT	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
RL	[3]			
RN	SEQUENCE FROM N.A.			
RP	MEDLINE=20510030; PubMed=11054574;			
RC	Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,			
RA	Moss P., Pieper B., Wang K.;			
RA	"Sequencing and expression analysis of the serine protease gene			
RT	cluster located in chromosome 19q13 region.";			
RT	Gene 257:119-130(2000).			
RL	[4]			
RN	CHARACTERIZATION.			
RP	MEDLINE=95314630; PubMed=7794273;			
RC	Skytt A., Stroemqvist M., Egelrud T.;			
RA	"Primary substrate specificity of recombinant human stratum corneum			
RT	chymotryptic enzyme.";			
RT	Biochem. Biophys. Res. Commun. 211:586-589(1995).			
RL	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC	CC			
CC				

CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; L33404; AAC37551.1; -

DR EMBL; AF166330; AAD49718.1; -

DR EMBL; AF243527; AAG33360.1; -

DR HSSP; P00763; IDPO. -

DR MEROPS; S01.300; -

DR MIN; 604438; -

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; trypsin.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; TRYP-SPC; 1.

DR PROSITE; PS00240; TRYPsin_DOM; 1.

DR PROSITE; PS00134; TRYPsin_HIS; 1.

DR PROSITE; PS00135; TRYPsin_SER; 1.

KW Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal.

FT SIGNAL 1 22

FT PROPEP 23 29

FT CHAIN 30 253

FT ACT_SITE 70 70

FT ACT_SITE 112 112

FT ACT_SITE 205 205

FT DISULFID 36 137

FT DISULFID 55 71

FT DISULFID 137 239

FT DISULFID 144 211

FT DISULFID 176 190

FT DISULFID 201 226

FT CARBOHYD 246 246

SQ SEQUENCE 253 AA; 27525 MW; 2D60B6B15A76A668 CRC64;

Query Match 100.0%; Score 41; DB 1; Length 253;

Best Local Similarity 100.0%; Pred. No. 0.16;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9

|||||||

Db 1 MARSLLLPL 9

RESULT 2

BAE2_HUMAN STANDARD; PRT; 518 AA.

AC Q9Y520; Q9YJ76;

DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Beta secretase 2 precursor (EC 3.4.23.-) (Beta-site APP-cleaving

DE enzyme 2) (Aspartyl protease 1) (Asp 1) (ASPI) (Membrane-associated

DE aspartic protease 1) (Memapsin-1).

GN BACE2 OR ASP21

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=20057170; PubMed=10591213;

RA Yan R., Bienkowski M.J., Shuck M.E., Miao H., Tory M.C., Pauley A.M.,

RA Brashier J.R., Stratton N.C., Mathews W.R., Buhl A.E., Carter D.B.,

RA Tomasselli A.G., Parodi L.A., Heinrikson R.L., Gurney M.E.;

RT "Membrane-anchored aspartyl protease with Alzheimer's disease

RT beta-secretase activity."

RL Nature 402:533-537(1999).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Bone marrow;

RA Xin H., Stephens J.C., Duan X., Harrowe G., Kim E., Grieshammer U.,

RA Giese K.;

RT "Identification of a novel aspartic-like protease differentially

RT expressed in human breast cancer cell lines."

RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RA Accarino M.P., Fumagalli P., Ottolenghi S., Taramelli R.;

RT "Cloning of a gene from chromosome 21 Down region encoding a potential

RT transmembrane aspartyl protease."

RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.

RN [4]

RP SEQUENCE FROM N.A.

RA Solans A., Estivill X., de la Luna S.;

RT "Cloning of a novel mammalian aspartyl protease."

RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE=20030166; PubMed=10561122;

RA Hussain I., Powell D.J., Howlett D.R., Tew D.G., Meek T.D.,

RA Chapman C., Gloger I.S., Murphy K.E., Southan C.D., Ryan D.M.,

RA Smith T.S., Simmons D.L., Walsh F.S., Dingwall C., Christie G.;

RT "Identification of a novel aspartic proteinase (Asp 2) as

RT beta-secretase."

RL Mol. Cell. Neurosci. 14:419-427(1999).

RN [6]

RP SEQUENCE FROM N.A.

RX MEDLINE=20144060; PubMed=10677483;

RA Lin X., Koelsch G., Wu S., Downs D., Dashti A., Tang J.;

RT "Human aspartic protease memapsin 2 cleaves the beta-secretase site of

RT beta-amyloid precursor protein."

RL Proc. Natl. Acad. Sci. U.S.A. 97:1456-1460(2000).

RN [7]

RP SEQUENCE FROM N.A.

RX MEDLINE=20289799; PubMed=10830953;

RA Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,

RA Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Soeda E.,

RA Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K., Polley A.,

RA Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,

RA Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,

RA Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,

RA Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,

RA Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,

RA Scharfe M., Schoen O., Desario A., Reicheit J., Kauer G., Bloecker H.,

RA Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,

RA Wehmeyer S., Borzym K., Gardiner K., Nizetic D., Francis F.,

RA Lehrach H., Reinhardt R., Yaspo M.-L.;

RT "The DNA sequence of human chromosome 21."

RL Nature 405:311-319(2000).

CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY A1; ALSO KNOWN AS THE

CC EUKARYOTIC ASPARTYL PROTEASES FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AF200342; AAF17078.1; -

DR EMBL; AF117892; AAD45240.1; -

DR EMBL; AF050171; AAD45963.1; -

DR EMBL; AF178532; AAF29494.1; -

DR EMBL; AF204944; AAF26368.1; -

DR EMBL; AF200192; AAF13714.1; -

DR EMBL; AL163284; CAB90458.1; -

DR EMBL; AL163285; CAB90554.1; -

DR HSSP; P00797; 2REN.

```
DR MIM; 605668; .
DR InterPro; IPR001969; Asp_protease.
DR Pfam; PF00026; asp; 3
DR PRINTS; PR00792; PEPsin.
DR PROSITE; PS00141; ASP_PROTEASE; 2.
KW Hydrolase; Aspartyl protease; Glycoprotein; Zymogen; Transmembrane;
Signal.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 ?
FT CHAIN ? 518 BETA SECRETASE 2.
FT DOMAIN 21 473 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 474 494 POTENTIAL.
FT DOMAIN 495 518 CYTOPLASMIC (POTENTIAL).
FT ACT_SITE 110 110 BY SIMILARITY.
FT ACT_SITE 303 303 BY SIMILARITY.
FT CARBOHYD 170 170 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CONFLICT 36 36 A -> T (IN REF. 6).
SQ SEQUENCE 518 AA; 56180 MW; 2E903150823760D3 CRC64;

Query Match 85.4%; Score 35; DB 1; Length 518;
Best Local Similarity 77.8%; Pred. No. 6.2;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
DB 4 LARALLPL 12
[1]

RESULT 3
ID SSL1 YEAST STANDARD; PRT; 461 AA.
AC Q04673;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Suppressor of stem-loop protein 1.
GN SSL1 OR YLR005W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94040711; PubMed=1340463;
RA Yoon H., Miller S.P., Pabich E.K., Donahue T.F.;
RT "SSL1, a suppressor of a His4 5'-UTR stem-loop mutation, is essential
RT for translation initiation and affects UV resistance in yeast.";
RL Genes Dev. 6:2463-2477(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Vandenbol M., Portetelle D., Hilger F.;
RL Submitted (MAY 1996) to the EMBL/GenBank/DDBJ databases.
RN [3]
RP SEQUENCE OF 400-461 FROM N.A.
RX MEDLINE=94239498; PubMed=8183345;
RA Maeda T., Wurgler-Murphy S.M., Saito H.;
RT "A two-component system that regulates an osmosensing MAP kinase
RT cascade in yeast.";
RL Nature 369:242-245(1994).
CC -1- FUNCTION: ESSENTIAL FOR TRANSLATION INITIATION AND AFFECTS
CC UV-RESISTANCE IN YEAST.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF000090; AAB91813.1;
DR InterPro; IPR001830; Glyco_transf_20.
DR Pfam; PF00982; Glyco_transf_20; 1.
KW Transferase; Glycosyltransferase; Plasmid.
SQ SEQUENCE 464 AA; 51627 MW; 96DC610DD739FA73 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 464;
Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
```

```
DR EMBL; Z17385; CAA78992.1; .
DR EMBL; Z73177; CAA97527.1; .
DR EMBL; L26523; AAA35101.1; .
DR PIR; A46394; A46394.
DR TRANSFAC; T02191; .
DR SGD; S0003995; SSL1.
DR InterPro; IPR000822; Znf-C2H2.
DR Pfam; PF00096; zf-C2H2; 1.
DR SMART; SM00327; ZNF_C2H2; 1.
DR SMART; SM00355; ZNF_C2H2; 1.
DR PROSITE; PS00028; ZINC_FINGER_C2H2_1; 1.
DR PROSITE; PS00157; ZINC_FINGER_C2H2_2; FALSE_NEG.
KW zinc-finger; Nuclear protein; Metal-binding.
FT ZN_FING 427 449 C2H2-TYPE.
SQ SEQUENCE 461 AA; 52290 MW; 57ADC630B790B4F CRC64;

Query Match 82.9%; Score 34; DB 1; Length 461;
Best Local Similarity 77.8%; Pred. No. 9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
DB 216 MARGLLLPV 224
[1]

RESULT 4
ID OTSA_RHISN STANDARD; PRT; 464 AA.
AC P55612;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable alpha, alpha-trehalose-phosphate synthase (UDP-forming)
DE (EC 2.4.1.15) (Trehalose 6-phosphate synthase) (UDP-glucose-
DE glucosephosphate glucosyltransferase).
GN OTSA OR Y4PC.
OS Rhizobium sp. (strain NGR234).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=394;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97305956; PubMed=9163424;
RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
RA Perret A.;
RT "Molecular basis of symbiosis between Rhizobium and legumes.";
RL Nature 387:394-401(1997).
CC -1- CATALYTIC ACTIVITY: UDP-glucose + D-glucose 6-phosphate = UDP +
CC alpha, alpha-trehalose 6-phosphate.
CC -1- SIMILARITY: TO THE E.COLI (OTSA) AND YEAST ENZYME (TPS1/LIF1).
CC -1- SIMILARITY: STRONG, TO A NON-FUNCTIONAL COPY IN NGR234, FQ1-FQ2,
CC TRUNCATED BY A ISRM3-LIKE INSERTION ELEMENT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE000090; AAB91813.1;
DR InterPro; IPR001830; Glyco_transf_20.
DR Pfam; PF00982; Glyco_transf_20; 1.
KW Transferase; Glycosyltransferase; Plasmid.
SQ SEQUENCE 464 AA; 51627 MW; 96DC610DD739FA73 CRC64;

Query Match 80.5%; Score 33; DB 1; Length 464;
Best Local Similarity 77.8%; Pred. No. 15;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
```

```
Db 418 MARSLMPL 426
|||||:|
RESULT 5
MM14_MOUSE STANDARD; PRT: 582 AA.
AC P53690; O08645; O35369;
DT 01-OCT-1996 (Rel. 34, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Matrix metalloproteinase-14 precursor (EC 3.4.24.-) (MMP-14)
DE (Membrane-type matrix metalloproteinase 1) (MT-MMP 1) (MTMMP1)
DE (Membrane-type-1 matrix metalloproteinase) (MT1-MMP) (MTIMMP) (MMP-
DE X1) (MT-MMP).
DE MMP14 OR MTMMP.
GN Mus musculus (Mouse).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95224014; PubMed=7708715;
RA Okada A., Bellocq J.P., Rouyer N., Chenard M.P., Rio M.C.,
RA Chambon P., Basset P.;
RT "Membrane-type matrix metalloproteinase (MT-MMP) gene is expressed in
RT stromal cells of human colon, breast, and head and neck carcinomas.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:2730-2734(1995).
RN [2]
RP REVISIONS.
RA Odaka A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97467335; PubMed=9325265;
RA Apte S.S., Fukai N., Beier D.R., Olsen B.R.;
RT "The matrix metalloproteinase-14 (MMP-14) gene is structurally
RT distinct from other MMP genes and is co-expressed with the TIMP-2 gene
RT during mouse embryogenesis.";
RL J. Biol. Chem. 272:25511-25517(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=CD-1; TISSUE=Kidney;
RX MEDLINE=98311877; PubMed=9648071;
RA Ota K., Stetler-Stevenson W.G., Yang Q., Kumar A., Wada J.,
RA Kashiwara N., Wallner E.I., Kanwar Y.S.;
RT "Cloning of murine membrane-type-1-matrix metalloproteinase (MT-1-MMP)
RT and its metanephric developmental regulation with respect to MMP-2
RT and its inhibitor.";
RL Kidney Int. 54:131-142(1998).
RN [5]
RP FUNCTION
RX MEDLINE=99449306; PubMed=10520996;
RA Holmbeck K., Bianco P., Caterina J., Yamada S., Kromer M.,
RA Kuznetsov S.A., Mankani M., Robey P.G., Poole A.R., Pidoux I.,
RA Ward J.M., Birkedal-Hansen H.;
RT "MT1-MMP-deficient mice develop dwarfism, osteopenia, arthritis, and
RT connective tissue disease due to inadequate collagen turnover.";
RL Cell 99:81-92(1999).
CC -|- FUNCTION: ENDOPEPTIDASE THAT DEGRADES VARIOUS COMPONENTS OF THE
CC EXTRACELLULAR MATRIX, SUCH AS COLLAGEN. ACTIVATES PROGELATINASE A.
CC ESSENTIAL FOR PERICELLULAR COLLAGENOLYSIS AND MODELING OF SKELETAL
CC AND EXTRASKELETAL CONNECTIVE TISSUES DURING DEVELOPMENT.
CC -|- COPACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Type I membrane protein (potential).
CC -|- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN PLACENTA, KIDNEY, HEART,
CC LUNG, EMBRYONIC SKELETAL AND PERISKELETAL TISSUES.
CC -|- DEVELOPMENTAL STAGE: NOT DETECTED BEFORE DAY 10.5. AT DAY 12.5,
CC PROMINENTLY EXPRESSED IN LARGE ARTERIES AND THE UMBILICAL
CC ARTERIES, EXPRESSED AT LOWER LEVELS IN THE MYOCARDIUM,
CC CRANIOFACIAL MESENCHYME, NASAL EPITHELIUM AND LIVER CAPSULE. AT
CC DAYS 14.5 AND 17.5, EXPRESSED IN THE MUSCULOSKELETAL SYSTEM, AND
```

```
QY 2 ARSLLLPL 9
DB 8 SRSLLLPL 15

RESULT 6
MM14_RAT
ID MM14_RAT STANDARD; PRT; 582 AA.
AC O10739;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Matrix metalloproteinase-14 precursor (EC 3.4.24.-) (MMP-14)
DE (Membrane-type matrix metalloproteinase 1) (MT-MMP 1) (MTMMP1)
DE (Membrane-type-1 matrix metalloproteinase) (MT1-MMP) (MT1MMP) (MT-
MMP).
DE MMP14 OR MTMMP.
GN Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95224014; PubMed=7708715;
RA Okada A., Bellocq J.P., Rouyer N., Chenard M.P., Rio M.C.,
RA Chambon P., Basset P.;
RT "Membrane-type matrix metalloproteinase (MT-MMP) gene is expressed in
RT stromal cells of human colon, breast, and head and neck carcinomas.";
RL Proc. Natl. Acad. Sci. U.S.A. 92:2730-2734(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Cossins J., Clements J., Catlin G., Wells G.;
RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: SEEMS TO SPECIFICALLY ACTIVATE OF PRO-GLATINASE A. MAY
CC THIS TRIGGER INVASION BY TUMOR CELLS BY ACTIVATING PRO-GLATINASE
CC A ON THE TUMOR CELL SURFACE.
CC -!- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential)..
CC -!- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
CC METALLOPROTEASE) ALSO KNOWN AS MATRIXIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X83537; CAA58521.1; -;
DR EMBL; X91785; CAA62897.1; -;
DR HSSP; P08254; IUSN.
DR MEROPS; M10.014; -.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001818; Matrixin.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00045; hemopexin; 4.
DR Pfam; PF00413; Peptidase_M10; 1.
DR PRINTS; PR00138; MATRINXIN.
DR SMART; SM00120; HX; 4.
DR SMART; SM00235; ZnMC; 1.
DR PROSITE; PS00024; HEMOPEXIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
KW Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium; Signal;
Transmembrane.
FT SIGNAL 1 20 POTENTIAL.
FT PROPEP 21 111 ACTIVATION PEPTIDE.
FT CHAIN 112 582 MATRIX METALLOPROTEINASE-14.
FT DOMAIN 112 541 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 542 562 POTENTIAL.
```

```
FT DOMAIN 563 582 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 316 511 HEMOPEXIN-LIKE.
FT SITE 93 CYSTEINE SWITCH (POTENTIAL).
FT METAL 239 239 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 240 240 BY SIMILARITY.
FT METAL 243 243 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 249 249 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 319 508 BY SIMILARITY.
FT CONFLICT 68 68 I -> M (IN REF. 2).
FT CONFLICT 255 255 D -> A (IN REF. 2).
SQ SEQUENCE 582 AA; 66106 MW; 8B40FDD9999CA80C CRC64;

Query Match 80.5%; Score 33; DB 1; Length 582;
Best Local Similarity 87.5%; Pred. No. 18;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
DB 8 SRSLLLPL 15

RESULT 7
MM14_PIG
ID MM14_PIG STANDARD; PRT; 580 AA.
AC Q9XTG0;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Matrix metalloproteinase-14 precursor (EC 3.4.24.-) (MMP-14)
DE (Membrane-type matrix metalloproteinase 1) (MT-MMP 1) (MTMMP1)
DE (Membrane-type-1 matrix metalloproteinase) (MT1-MMP) (MT1MMP).
DE MMP14.
GN Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99095929; PubMed=9881602;
RA Caron C., Xue J., Bartlett J.D.;
RT "Expression and localization of membrane type 1 matrix
RT metalloproteinase in tooth tissues.";
RL Matrix Biol. 17:501-511(1998).
CC -!- FUNCTION: SEEMS TO SPECIFICALLY ACTIVATE OF PRO-GLATINASE A. MAY
CC THIS TRIGGER INVASION BY TUMOR CELLS BY ACTIVATING PRO-GLATINASE
CC A ON THE TUMOR CELL SURFACE (BY SIMILARITY). MAY PLAY A ROLE IN
CC THE BIOMINERALIZATION OF ENAMEL AND DENTIN.
CC -!- COFACTOR: REQUIRES CALCIUM AND ZINC FOR ACTIVITY (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -!- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN DEVELOPING TOOTH TISSUES
CC AND MAY PLAY A ROLE IN THE BIOMINERALIZATION OF ENAMEL AND DENTIN.
CC -!- SIMILARITY: CONTAINS 1 HEMOPEXIN-LIKE DOMAIN.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M10A (ZINC
CC METALLOPROTEASE) ALSO KNOWN AS MATRIXIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF067419; AAD38324.1; -;
DR HSSP; P08254; IUSN.
DR MEROPS; M10.014; -.
DR InterPro; IPR000585; Hemopexin.
DR InterPro; IPR001818; Matrixin.
DR InterPro; IPR000130; Zn_MTpeptdse.
DR Pfam; PF00045; hemopexin; 4.
DR Pfam; PF00413; Peptidase_M10; 1.
DR PRINTS; PR00138; MATRINXIN.
DR SMART; SM00120; HX; 4.
```

```

DR SMART; SM00235; ZnMC; 1.
DR PROSITE; PS00024; HEMOPEXIN; 1.
DR PROSITE; PS00142; ZINC_PROTEASE; 1.
DR PROSITE; PS00546; CYSTEINE_SWITCH; 1.
KW Hydrolase; Metalloprotease; Zinc; Zymogen; Calcium; Signal;
KW Transmembrane.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 109 ACTIVATION PEPTIDE.
FT CHAIN 110 580 MATRIX METALLOPROTEINASE-14.
FT DOMAIN 110 539 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 540 560 POTENTIAL.
FT DOMAIN 561 580 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 314 509 HEMOPEXIN-LIKE.
FT SITE 91 91 CYSTEINE SWITCH (POTENTIAL).
FT METAL 237 237 ZINC (CATALYTIC) (BY SIMILARITY).
FT ACT_SITE 238 238 BY SIMILARITY.
FT METAL 241 241 ZINC (CATALYTIC) (BY SIMILARITY).
FT METAL 247 247 ZINC (CATALYTIC) (BY SIMILARITY).
FT DISULFID 317 506 BY SIMILARITY.
SQ SEQUENCE 580 AA; 65934 MW; B7B2C2C569A96CAC CRC64;

Query Match 78.0%; Score 32; DB 1; Length 580;
Best Local Similarity 100.0%; Pred. No. 30;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLLPL 9
DB 9 RSLLLPL 15

RESULT 8
SNXD_HUMAN STANDARD; PRT; 1025 AA.
AC Q9Y5W8; O94821;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sorting nexin 13 (Fragment).
GN SNX13 OR KIAA0713.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=99087487; PubMed=9872452;
RA Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Miyajima N.,
RA Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XI.
RT The complete sequences of 100 new cDNA clones from brain which code
RT for large proteins in vitro.";
RL DNA Res. 5:277-286(1998).
RN [2]
RP SEQUENCE OF 664-1025 FROM N.A.
RA Teasdale R.D., Gleeson P.A., Karlsson L.;
RT "Identification of eleven novel human sorting nexin molecules. A sub-
RT group of the sorting nexin family is associated with the early
RT endosomes.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: MAY BE INVOLVED IN SEVERAL STAGES OF INTRACELLULAR
CC TRAFFICKING.
CC -!- SIMILARITY: BELONGS TO THE SORTING NEXIN FAMILY.
CC -!- SIMILARITY: CONTAINS 1 RGS DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PHOX HOMOLOG (PX) DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 PXA DOMAIN.
CC -!- CAUTION: REF.1 SEQUENCE DIFFERS FROM THAT SHOWN DUE A FRAMESHIFT
CC IN POSITION 944.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z69239; CAA93223.1; -
DR EMBL; Z54285; CAA91079.2; -
KW Hypothetical protein.
SQ SEQUENCE 1628 AA; 188831 MW; 85A30FB43D2CAED0 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 1628;

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; AB018256; BAA34433.1; ALT_FRAME.
DR InterPro; IPR001683; PX.
DR InterPro; IPR003114; PX_assoc.
DR Pfam; PF00787; PX; 1.
DR Pfam; PF02194; PXA; 1.
DR Pfam; PF00615; RGS; 1.
DR SMART; SM00312; PX; 1.
DR SMART; SM00313; PXA; 1.
DR SMART; SM00315; RGS; 1.
DR PROSITE; PS0132; RGS; 1.
KW Transport; Protein transport.
FT NON_TER 1 1
FT DOMAIN 154 341 PXA.
FT DOMAIN 430 553 RGS.
FT DOMAIN 648 748 PX.
FT CONFLICT 695 695 E -> G (IN REF. 2).
SQ SEQUENCE 1025 AA; 118090 MW; 1748715CB24F2791 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 1025;
Best Local Similarity 66.7%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
DB 311 LARGILLPL 319

RESULT 9
YA8E_SCHPO STANDARD; PRT; 1628 AA.
AC Q09779; O13884;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 188.8 kDa protein C22F3.14c in chromosome 1.
GN SPAC22F3.14C OR SPAC1D4.14.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE OF 1-26 FROM N.A.
RC STRAIN=972;
RA Lye G., Churcher C.M.;
RL Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 8-1628 FROM N.A.
RC STRAIN=972;
RA Lye G., Churcher C.M., Barrell B.G., Rajandream M.A., Walsh S.V.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: TO YEAST RL1.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; Z69239; CAA93223.1; -
DR EMBL; Z54285; CAA91079.2; -
KW Hypothetical protein.
SQ SEQUENCE 1628 AA; 188831 MW; 85A30FB43D2CAED0 CRC64;

Query Match 78.0%; Score 32; DB 1; Length 1628;

```



```
Best Local Similarity 100.0%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLLPL 9
Db 583 RSLLLPL 589

RESULT 10
SULA_SALTY
ID SULA_SALTY STANDARD; PRT; 169 AA.
AC P08847;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Cell division inhibitor
GN SULA OR SPW1071 OR STY1092.
GE Salmonella typhimurium, and
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602, 601;
RN [1]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium;
RX MEDLINE=87248093; PubMed=3297925;
RA Freudl R., Braun G., Honore N., Cole S.T.;
RT "Evolution of the enterobacterial sula gene: a component of the SOS
RT system encoding an inhibitor of cell division.";
RL Gene 52:31-40(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RA "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE FROM N.A.
RC SPECIES=S.typhi; STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608;
RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,
RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebaihia M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,
RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
RA Whitehead S., Barrall B.G.;
RA "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
CC -!- FUNCTION: SULA IS A COMPONENT OF THE SOS SYSTEM AND AN INHIBITOR
CC OF CELL DIVISION. ACCUMULATION OF SULA CAUSES RAPID CESSATION OF
CC CELL DIVISION AND THE APPEARANCE OF LONG, NON-SEPTATE FILAMENTS.
CC THE EXPRESSION OF SULA IS REPPRESSED BY LEXA PROTEIN. FTSZ SEEMS
CC TO BE THE TARGET OF SULA.
CC -!- SUBCELLULAR LOCATION: INNER MEMBRANE.
CC -!- MISCELLANEOUS: THE N-TERMINUS OF SULA PROTEIN MIGHT BE INVOLVED IN
CC RECOGNIZING THE CELL DIVISION APPARATUS.
CC -!- SIMILARITY: THE C-TERMINUS SHOWS SIMILARITY TO THE N PROTEIN OF
CC PHAGE LAMBDA.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X97337; CAA66015.1; -
CC InterPro; IPR003917; NADHub_oxdrdctse2.
CC InterPro; IPR001750; Oxidored_q1.
CC Pfam; PF00361; oxidored_q1; 1.
CC PRINTS; PR01436; NADHDHGNASE2.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
CC SEQUENCE 346 AA; 38870 MW; 021D6D976DB564DB CRC64;

Query Match 75.6%; Score 31; DB 1; Length 346;
Best Local Similarity 77.8%; Pred. No. 29;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M16324; AAA27230.1; -
CC EMBL; AE008746; AAL20004.1; -
CC EMBL; AL627269; CAD08197.1; -
CC PIR; B29016; B29016.
CC StyGene; SG10386; sula.
CC Cell division; Septation; SOS response; Inner membrane;
KW Complete proteome.
FT DOMAIN 31 149 CONSERVED REGION.
FT SIMILAR 150 169 TO N PROTEIN OF PHAGE LAMBDA.
FT DOMAIN 160 169 LON PROTEIN BINDING SITE (PROBABLE).
FT CONFLICT 102 102 S -> T (IN REF. 1).
FT CONFLICT 108 108 R -> A (IN REF. 1).
FT CONFLICT 151 151 H -> L (IN REF. 1).
SQ SEQUENCE 169 AA; 19013 MW; 3848A73595ESD176 CRC64;

Query Match 75.6%; Score 31; DB 1; Length 169;
Best Local Similarity 77.8%; Pred. No. 14;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 45 MAQLLLPL 53

RESULT 11
NU2M_EQUAS
ID NU2M_EQUAS STANDARD; PRT; 346 AA.
AC P92476;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1999 (Rel. 36, Last annotation update)
DE NADH-ubiquinone oxidoreductase chain 2 (EC 1.6.5.3).
GN MTND2 OR ND2 OR NADH2.
OS Equus asinus (Donkey).
OG Mitochondrion.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9793;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RX MEDLINE=97032591; PubMed=8875857;
RA Xu X., Gullberg A., Arnason U.;
RT "The complete mitochondrial DNA (mtDNA) of the donkey and mtDNA
RT comparisons among four closely related mammalian species-pairs.";
RL J. Mol. Evol. 43:438-463(1996).
CC -!- CATALYTIC ACTIVITY: NADH + ubiquinone = NAD(+) + ubiquinol.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein. Mitochondrial
CC inner membrane.
CC -!- SIMILARITY: BELONGS TO THE COMPLEX I SUBUNIT 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X97337; CAA66015.1; -
CC InterPro; IPR003917; NADHub_oxdrdctse2.
CC InterPro; IPR001750; Oxidored_q1.
CC Pfam; PF00361; oxidored_q1; 1.
CC PRINTS; PR01436; NADHDHGNASE2.
CC Oxidoreductase; NAD; Ubiquinone; Mitochondrion; Transmembrane.
CC SEQUENCE 346 AA; 38870 MW; 021D6D976DB564DB CRC64;

Query Match 75.6%; Score 31; DB 1; Length 346;
Best Local Similarity 77.8%; Pred. No. 29;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

Qy 1 MARSLLLPL 9
: | | | | |
Db 330 IASSLLPL 338

RESULT 12

CIN2_RAT STANDARD; PRT; 2005 AA.
AC P04775;
DT 13-AUG-1987 (Rel. 05, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Sodium channel protein, brain II alpha subunit.
GN SCN2A1 OR SCN2A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86146901; PubMed=3754035;
RA Noda M., Ikeda T., Kayano T., Suzuki H., Takeshima H., Kurasaki M.,
RA Takahashi H., Numa S.;
RT "Existence of distinct sodium channel messenger RNAs in rat brain.";
RL Nature 320:188-192(1986).
CC -1- FUNCTION: THIS PROTEIN MEDIATES THE VOLTAGE-DEPENDENT SODIUM ION
CC PERMEABILITY OF EXCITABLE MEMBRANES. ASSUMING OPENED OR CLOSED
CC CONFORMATIONS IN RESPONSE TO THE VOLTAGE DIFFERENCE ACROSS THE
CC MEMBRANE, THE PROTEIN FORMS A SODIUM-SELECTIVE CHANNEL THROUGH
CC WHICH NA++ IONS MAY PASS IN ACCORDANCE WITH THEIR ELECTROCHEMICAL
CC GRADIENT.
CC -1- SUBUNIT: THE SODIUM CHANNEL CONSISTS OF A LARGE POLYPEPTIDE AND
CC 2-3 SMALLER ONES. THIS SEQUENCE REPRESENTS A LARGE POLYPEPTIDE.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- DOMAIN: THE SEQUENCE CONTAINS 4 INTERNAL REPEATS, EACH WITH 5
CC HYDROPHOBIC SEGMENTS (S1,S2,S3,S5,S6) AND ONE POSITIVELY CHARGED
CC SEGMENT (S4). SEGMENTS S4 ARE PROBABLY THE VOLTAGE-SENSORS AND ARE
CC CHARACTERIZED BY A SERIES OF POSITIVELY CHARGED AMINO ACIDS AT
CC EVERY THIRD POSITION.
CC -1- SIMILARITY: TO OTHER SODIUM CHANNEL PROTEINS.
CC -1- SIMILARITY: CONTAINS 1 IQ DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X03639; CAA27287.1; -.
DR PIR; B25019; B25019.
DR InterPro; IPR002111; Cat_channel_TrpL.
DR InterPro; IPR000636; Cation_chan_non_lig.
DR InterPro; IPR001682; Channel_pore_Ca_Na.
DR InterPro; IPR000048; IQ.
DR InterPro; IPR001696; Na_channel.
DR Pfam; PF00520; ion_trans; 4.
DR PRINTS; PF00612; IQ_1.
DR SMART; SM00015; IQ; 1.
DR PROSITE; PS00096; IQ; 1.
DR Ionic channel; Transmembrane; Ion transport; Voltage-gated channel;
KW Glycoprotein; Repeat; Multigene family.
FT REPEAT 111 456 I.
FT REPEAT 741 1013 II.
FT REPEAT 1190 1504 III.
FT REPEAT 1513 1811 IV.
FT TRANSMEM 125 148 S1 OF REPEAT I.
FT TRANSMEM 157 176 S2 OF REPEAT I.
FT TRANSMEM 190 208 S3 OF REPEAT I.
FT TRANSMEM 215 234 S4 OF REPEAT I.

FT TRANSMEM 251 274 S5 OF REPEAT I.
FT TRANSMEM 402 427 S6 OF REPEAT I.
FT TRANSMEM 754 778 S1 OF REPEAT II.
FT TRANSMEM 790 813 S2 OF REPEAT II.
FT TRANSMEM 822 841 S3 OF REPEAT II.
FT TRANSMEM 848 867 S4 OF REPEAT II.
FT TRANSMEM 884 904 S5 OF REPEAT II.
FT TRANSMEM 958 983 S6 OF REPEAT II.
FT TRANSMEM 1204 1227 S1 OF REPEAT III.
FT TRANSMEM 1241 1266 S2 OF REPEAT III.
FT TRANSMEM 1273 1294 S3 OF REPEAT III.
FT TRANSMEM 1299 1320 S4 OF REPEAT III.
FT TRANSMEM 1340 1367 S5 OF REPEAT III.
FT TRANSMEM 1447 1473 S6 OF REPEAT III.
FT TRANSMEM 1527 1550 S1 OF REPEAT IV.
FT TRANSMEM 1562 1585 S2 OF REPEAT IV.
FT TRANSMEM 1592 1615 S3 OF REPEAT IV.
FT TRANSMEM 1626 1647 S4 OF REPEAT IV.
FT TRANSMEM 1663 1685 S5 OF REPEAT IV.
FT TRANSMEM 1752 1776 S6 OF REPEAT IV.
FT DOMAIN 1905 1934 IQ.
FT CARBOHYD 212 212 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 285 285 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 291 291 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 297 297 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 303 303 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 308 308 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 340 340 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 604 604 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 624 624 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 883 883 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1055 1055 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1072 1072 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1136 1136 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1368 1368 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1382 1382 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1393 1393 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 2005 AA; 227872 MW; 861BE583D79F8324 CRC64;

Query Match 75.68; Score 31; DB 1; Length 2005;

Best Local Similarity 75.08; Pred. No. 1.7e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLP 8

Db 1 MARSVLVP 8

RESULT 13

CORZ_DROME STANDARD; PRT; 72 AA.
AC Q26377; Q9VFK7;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Crz precursor [Contains: Corazonin; Corazonin-precursor-related
DE peptide (CPRP)].
GN CRZ OR CG3302.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95032110; PubMed=7945373;
RA Veenstra J.A.;
RT "Isolation and structure of the Drosophila corazonin gene.";
RL Biochem. Biophys. Res. Commun. 204:292-296(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon K.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balaw R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissensbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 CC -!- FUNCTION: CARDIOACTIVE PEPTIDE. CORAZONIN IS PROBABLY INVOLVED
 CC IN THE PHYSIOLOGICAL REGULATION OF THE HEART BEAT.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: S74038; AAB32283.1; -.
 CC DR EMBL: AE003704; AAF55046.1; ALT_SEQ.
 CC DR FlyBase: FBgn013767; Crz.
 KW Neuropeptide; Amidation; Cleavage on pair of basic residues; Signal.
 FT SIGNAL 1 19
 FT PEPTIDE 20 30 CORAZONIN.
 FT PEPTIDE 34 67 CORAZONIN-PRECURSOR-RELATED PEPTIDE.
 FT PROPEP 68 72
 FT MOD_RES 20 20 PYRROLIDONE CARBOXYLIC ACID.
 FT MOD_RES 30 30 AMIDATION (G-31 PROVIDE AMIDE GROUP).
 SQ SEQUENCE 72 AA: 8331 MW; 3676CCCCF8D5107E CRC64;
 Query Match 73.2%; Score 30; DB 1; Length 72;
 Best Local Similarity 77.8%; Pred. No. 10;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 OY 1 MARSLLLPL 9
 DB 1 MLRLLLPL 9
 RESULT 14
 CYTC_RABIT
 ID CVTC_RABIT STANDARD; PRT; 148 AA.

AC O97862;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 GN Cystatin C precursor.
 GN CST3.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OC NCBI_TaxID=9986;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=JAPAN WHITE; TISSUE=Bone;
 RX MEDLINE=98424349; PubMed=9753427;
 RA Kobori M., Ikeda Y., Nara H., Kato M., Kumegawa M., Nojima H.,
 RA Kawashima H.;
 RT "Large scale isolation of osteoclast-specific genes by an improved
 RT method involving the preparation of a subtracted cDNA library.";
 RL Genes Cells 3:459-475(1998).
 CC -!- FUNCTION: THIS IS A THIOL PROTEINASE INHIBITOR.
 CC -!- SIMILARITY: BELONGS TO THE CYSTATIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: AB009342; BAA75921.1; -.
 CC DR HSP: P01038; ICEW.
 DR InterPro: IPR000010; Cystatin.
 DR InterPro: IPR003243; Cystatin_C_M.
 DR Pfam: PF00031; cystatin; 1.
 DR ProDom: PD001231; Cystatin_C_M; 1.
 DR SMART: SM00043; CV; 1.
 DR PROSITE: PS00287; CYSTATIN; FALSE_NEG.
 KW Thiol protease inhibitor; Signal.
 FT SIGNAL 1 28 POTENTIAL.
 FT CHAIN 29 148 CYSTATIN C.
 FT ACT_SITE 39 39 REACTIVE SITE.
 FT SITE 83 87 SECONDARY AREA OF CONTACT.
 FT DISULFID 101 111 BY SIMILARITY.
 FT DISULFID 125 145 BY SIMILARITY.
 SQ SEQUENCE 148 AA: 16346 MW; 1523C8311695B9A CRC64;
 Query Match 73.2%; Score 30; DB 1; Length 148;
 Best Local Similarity 77.8%; Pred. No. 21;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 OY 1 MARSLLLPL 9
 DB 1 MARSLGVL 9
 RESULT 15
 YHJC_ECOLI
 ID YHJC_ECOLI STANDARD; PRT; 299 AA.
 AC P37641;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical transcriptional regulator yhjC.
 GN YHJC OR B3521.
 OS Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=94316500; PubMed=8041620;

RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RL region from 76.0 to 81.5 minutes.";
RC Nucleic Acids Res. 22:2576-2586(1994).
CC -1- SIMILARITY: BELONGS TO THE LYSR FAMILY OF TRANSCRIPTIONAL
CC REGULATORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U00039; AAB18497.1; ALT_INIT.
DR EMBL: AE000428; AAC76546.1; ALT_INIT.
DR EcoGene: EG12247; yhjC.
DR InterPro: IPR000847; HTH_LysR.
DR Pfam: PF00126; HTH_1; 1.
DR PROSITE: PS00044; HTH_LYSR_FAMILY; 1.
KW Hypothetical protein; Transcription regulation; DNA-binding;
KW Complete proteome.
FT DNA_BIND 19 38 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 299 AA; 33329 MW; 22D0EC5994CBC8C1 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 299;
Best Local Similarity 75.0%; Pred. No. 42;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLPL 8
Db 102 IAKSLPL 109

RESULT 16
CD68_HUMAN STANDARD; PRT; 354 AA.
ID CD68_HUMAN
AC P34810; Q96BI7;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Macroscialin precursor (CD68 antigen) (GP110).
GN CD68.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RS SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93200523; PubMed=7680921;
RA Holness C.L., Simmons D.L.;
RT "Molecular cloning of CD68, a human macrophage marker related to
RT lysosomal glycoproteins,"
RL Blood 81:1607-1613(1993).
RN [2]
RS SEQUENCE FROM N.A.
RC Tissue-Skin;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RS SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=99009345; PubMed=9790779;
RA Jones E., Quinn C.M., See C.G., Montgomery D.S., Ford M.J.,
RA Koelble K., Gordon S., Greaves D.R.;
RT "The linked human elongation initiation factor 4A1 (EIF4A1) and CD68
RT genes map to chromosome 17p13.";
RL Genomics 53:248-250(1998).
CC -1- FUNCTION: COULD PLAY A ROLE IN PHAGOCYTIC ACTIVITIES OF TISSUE
CC MACROPHAGES, BOTH IN INTRACELLULAR LYSOSOMAL METABOLISM AND
CC EXTRACELLULAR CELL-CELL AND CELL-PATHOGEN INTERACTIONS.
CC BIND TO TISSUE- AND ORGAN-SPECIFIC LECTINS OR SELECTINS, ALLOWING
CC HOMING OF MACROPHAGE SUBSETS TO PARTICULAR SITES. RAPID

CC RECIRCULATION OF CD68 FROM ENDOSOMES, LYSOSOMES TO THE PLASMA
CC MEMBRANE MAY ALLOW MACROPHAGES TO CRAWL OVER SELECTIN BEARING
CC SUBSTRATES OR OTHER CELLS.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOSOMAL OR
CC LYSOSOMAL (LONG VARIANT) AND TO A LESSER EXTENT ON THE CELL
CC SURFACE (SHORT VARIANT).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED BY BLOOD MONOCYTES AND
CC TISSUE MACROPHAGES. ALSO EXPRESSED IN MANY TUMOR CELL LINES WHICH
CC COULD ALLOW THEM TO ATTACH TO SELECTINS ON VASCULAR ENDOTHELIUM,
CC FACILITATING THEIR DISSEMINATION TO SECONDARY SITES.
CC -1- PTM: N- AND O-GLYCOSYLATED.
CC -1- SIMILARITY: BELONGS TO THE LAMP FAMILY.
CC -1- DATABASE: NAME=PROM; NOTE=CD guide CD68 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd68.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S57235; AAB25811.1; -
DR EMBL: BC015557; AAH15557.1; -
DR EMBL: AF060540; AAC70006.1; -
DR PIR: A48931; A48931.
DR MIM: 153634; -
DR InterPro: IPR002000; Lamp.
DR Pfam: PF01299; Lamp; 1.
DR PRINTS: PR00336; LYSASSOCTDMP.
DR PROSITE: PS00311; LAMP_2; 1.
KW Transmembrane; Glycoprotein; Signal; Lysosome; Repeat; Antigen;
KW Alternative splicing.
FT SIGNAL 1 21
FT CHAIN 22 354
FT DOMAIN 22 319
FT TRANSMEM 320 344
FT DOMAIN 345 354
FT DOMAIN 23 140
FT DOMAIN 140 152
FT DOMAIN 70 129
FT REPEAT 70 99
FT REPEAT 100 129
FT DISULFID 169 207
FT DISULFID 277 314
FT CARBOHYD 88 88
FT CARBOHYD 96 96
FT CARBOHYD 118 118
FT CARBOHYD 126 126
FT CARBOHYD 164 164
FT CARBOHYD 199 199
FT CARBOHYD 246 246
FT CARBOHYD 261 261
FT CARBOHYD 279 279
FT VARSPIC 17 43
FT VARSPIC 83 112
FT CONFLICT 254 254
SQ SEQUENCE 354 AA; 37408 MW; 0A29ACBBF9431B0F CRC64;

Query Match 73.2%; Score 30; DB 1; Length 354;
Best Local Similarity 85.7%; Pred. No. 49;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RSLLLPL 9
Db 318 RSLLLPL 324

RESULT 17
CYHY_G1BBA

```

ID CYHY_GIBBA STANDARD; PRT; 357 AA.
AC P32963;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Cyanide hydratase (EC 4.2.1.56) (Formamide hydrolyase).
GN CHY1.
OS Gibberella baccata (Fusarium lateritium).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OX NCBI_TaxID=5523;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=94014978; PubMed=8409923;
RA Clunness M.J., Turner P.D., Clements E., Brown D.T., O'Reilly C.;
RT "Purification and properties of cyanide hydratase from Fusarium
RT lateritium and analysis of the corresponding chl1 gene.";
RL J. Gen. Microbiol. 139:1807-1815(1993).
CC -|- FUNCTION: DETOXIFIES HCN AND IS THOUGHT TO BE IMPORTANT IN FUNGAL
CC INFECTION OF CYANOGENIC PLANTS.
CC -|- CATALYTIC ACTIVITY: Formamide = cyanide + H(2)O.
CC -|- SUBUNIT: MULTIMERIC.
CC -|- SIMILARITY: BELONGS TO THE NITRILASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M99046; AAA33336.1; -.
DR InterPro; IPR003010; CN_hydrolase.
DR InterPro; IPR000132; Nitril_cyn_hydratase.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
KW Lyase.
FT ACT_SITE 164 164 BY SIMILARITY.
SQ SEQUENCE 357 AA; 40353 MW; 761419C82744EA19 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 357;
Best Local Similarity 75.0%; Pred. No. 50;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLP 8
DB 222 LARGLLLP 229

RESULT 18
PTL_ECOLI
ID PTL_ECOLI STANDARD; PRT; 575 AA.
AC P08839;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9)
DE (Phosphotransferase system, enzyme I).
GN PTSI OR B2416.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
[1]
RN SEQUENCE FROM N.A.
RX MEDLINE=88314869; PubMed=2457575;
RA de Reuse H., Danchin A.;
RT "The ptsH, ptsI, and crr genes of the Escherichia coli
RT phosphoenolpyruvate-dependent phosphotransferase system: a complex
RT operon with several modes of transcription.";
RL J. Bacteriol. 170:3827-3837(1988).

```

```

RN SEQUENCE FROM N.A.
RX MEDLINE=88058992; PubMed=2960675;
RA Saffen D.W., Presper K.A., Doering T.L., Roseman S.;
RT "Sugar transport by the bacterial phosphotransferase system.
RT Molecular cloning and structural analysis of the Escherichia coli
RT ptsH, ptsI, and crr genes.";
RL J. Biol. Chem. 262:16241-16253(1987).
[3]
RN SEQUENCE FROM N.A.
RX STRAIN=K12 / MG1655;
RC MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[4]
RN SEQUENCE FROM N.A.
RX STRAIN=K12;
RC MEDLINE=97349980; PubMed=9205837;
RA Yamamoto Y., Alba H., Baba T., Hayashi K., Inada T., Isono K.,
RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,
RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nashimoto H.,
RA Oshima T., Oyama S., Saito N., Sampei G., Satoh Y., Sivasubaram S.,
RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,
RA Yamagata S., Horiuchi T.;
RT "Construction of a contiguous 874-kb sequence of the Escherichia coli
RT K-12 genome corresponding to 50.0-68.8 min on the linkage map and
RT analysis of its sequence features.";
RL DNA Res. 4:91-113(1997).
[5]
RN SEQUENCE OF 1-56 FROM N.A.
RX MEDLINE=85286351; PubMed=2411636;
RA de Reuse H., Roy A., Danchin A.;
RT "Analysis of the ptsH-ptsI-crr region in Escherichia coli K-12:
RT nucleotide sequence of the ptsH gene.";
RL Gene 35:199-207(1985).
[6]
RN SEQUENCE OF 1-54 FROM N.A.
RX STRAIN=K12;
RC MEDLINE=88257033; PubMed=3290198;
RA Byrne C.R., Monroe R.S., Ward K.A., Kredich N.M.;
RT "DNA sequences of the cysK regions of Salmonella typhimurium and
RT Escherichia coli and linkage of the cysK regions to ptsH.";
RL J. Bacteriol. 170:3150-3157(1988).
[7]
RN SEQUENCE OF 1-11.
RX STRAIN=K12 / EMG2;
RC MEDLINE=97443975; PubMed=9298646;
RA Link A.J., Robison K., Church G.M.;
RT "Comparing the predicted and observed properties of proteins encoded
RT in the genome of Escherichia coli K-12.";
RL Electrophoresis 18:1259-1313(1997).
[8]
RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF 1-259.
RX MEDLINE=96434331; PubMed=8805571;
RA Liao D.-I., Silverton E., Seok Y.-J., Lee B.R., Peterkofsky A.,
RA Davies D.R.;
RT "The first step in sugar transport: crystal structure of the amino
RT terminal domain of enzyme I of the E. coli PEP: sugar
RT phosphotransferase system and a model of the phosphotransfer complex
RT with HPr.";
RL Structure 4:861-872(1996).
[9]
RN STRUCTURE BY NMR OF 1-259.
RX MEDLINE=97207064; PubMed=9054557;
RA Garrett D.S., Seok Y.-J., Liao D.-I., Peterkofsky A., Gronenborn A.M.,
RA Clore G.M.;
RT "Solution structure of the 30 kDa N-terminal domain of enzyme I of
RT the Escherichia coli phosphoenolpyruvate:sugar phosphotransferase
RT system by multidimensional NMR.";

```

Biochemistry 36:2517-2530(1997).
[10]
RN STRUCTURE BY NMR OF 1-259.
RX MEDLINE=98200485; PubMed=9541412;
RA Garrett D.S., Seok Y.-J., Peterkofsky A., Clore G.M., Gronenborn A.M.;
RT "tautomeric state and pKa of the phosphorylated active site histidine
in the N-terminal domain of enzyme I of the Escherichia coli
phosphoenolpyruvate:sugar phosphotransferase system.";
RN Protein Sci. 7:789-793(1998).
[11]
RN STRUCTURE BY NMR OF 1-259.
RX MEDLINE=99140298; PubMed=10048929;
RA Garrett D.S., Seok Y.-J., Peterkofsky A., Gronenborn A.M., Clore G.M.;
RT "Solution structure of the 40,000 Mr phosphoryl transfer complex
between the N-terminal domain of enzyme I and Hpr.";
RN Nat. Struct. Biol. 6:166-173(1999).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
-TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM
PHOSPHOENOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).
CC ENZYME I IS COMMON TO ALL PTS.
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + protein L-histidine =
pyruvate + protein N(pi)-phospho-L-histidine.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: J02796; AAA24441.1; -;
DR EMBL: M10425; AAA24439.1; -;
DR EMBL: AE000329; AAC75469.1; -;
DR EMBL: D90871; BAA16290.1; -;
DR EMBL: M21994; AAA24385.1; -;
DR EMBL: M21451; AAA23656.1; -;
DR PIR: B29785; WOECPT.
DR PIR: B24035; B24035.
DR PIR: H28181; H28181.
DR PDB: 1EZA; 07-JAN-98.
DR PDB: 2EZA; 20-AUG-97.
DR PDB: 1EZB; 07-JAN-98.
DR PDB: 2EZB; 20-AUG-97.
DR PDB: 1EZC; 07-JAN-98.
DR PDB: 2EZC; 20-AUG-97.
DR PDB: 1EZD; 07-JAN-98.
DR PDB: 3EZB; 29-DEC-99.
DR PDB: 3EZE; 16-DEC-98.
DR PDB: 1ZYV; 07-DEC-96.
DR SWISS-2DPAGE; P08839; COLI.
DR ECO2DBASE; B058.3; 6TH EDITION.
DR EcoGene; EGI0789; ptsI.
DR InterPro; IPR000121; PEP-utilizers.
DR Pfam; PF00391; PEP-utilizers; 1.
DR Pfam; PF02896; PEP-utilizers_C; 1.
DR ProDom; PD000940; PEP-utilizers; 1.
DR PROSITE; PS00370; PEP-ENZYMES_PHOS_SITE; 1.
DR PROSITE; PS00742; PEP-ENZYMES_2; 1.
KW Phosphotransferase system; Transferase; Kinase; Sugar transport;
FT MOD_RES 189 189 PHOSPHORYLATION
SQ SEQUENCE 575 AA; 63561 MW; 4278f0838855950 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 575;
Best Local Similarity 87.5%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLP 8
||||| ||

||||| ||
DB 193 MARSLP 200

RESULT 19
PTI_HAEIN STANDARD; PRT; 575 AA.
ID AC P43922;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9)
DE {Phosphotransferase system, enzyme I}.
GN PTSI OR H11712.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.D., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
influenzae Rd.";
RL Science 269:496-512(1995).
CC -1- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT
SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE
-TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM
PHOSPHOENOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).
CC ENZYME I IS COMMON TO ALL PTS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Phosphoenolpyruvate + protein L-histidine =
pyruvate + protein N(pi)-phospho-L-histidine.
CC -1- SUBUNIT: HOMODIMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: U32844; AAC23357.1; -;
DR HSSP; P08839; LEZC.
DR TIGR; H11712; -;
DR InterPro; IPR000121; PEP-utilizers.
DR Pfam; PF00391; PEP-utilizers; 1.
DR Pfam; PF02896; PEP-utilizers_C; 1.
DR ProDom; PD000940; PEP-utilizers; 1.
DR PROSITE; PS00370; PEP-ENZYMES_PHOS_SITE; 1.
DR PROSITE; PS00742; PEP-ENZYMES_2; 1.
KW Phosphotransferase system; Transferase; Kinase; Sugar transport;
FT MOD_RES 189 189 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 575 AA; 63691 MW; DAC2B7909C38CE73 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 575;
Best Local Similarity 87.5%; Pred. No. 80;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLP 8
||||| ||

Db 193 MARSLELP 200

RESULT 20

PT1_SALTY

ID PT1_SALTY STANDARD; PRT; 575 AA.

AC P12654;

DT 01-OCT-1989 (Rel. 12, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Phosphoenolpyruvate-protein phosphotransferase (EC 2.7.3.9)

DE {Phosphotransferase system, enzyme I}.

GN PTSI OR STM2432.

OS Salmonella typhimurium.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Salmonella.

OX NCBI_TaxID=602;

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE=92011751; PubMed=1655788;

RA Licalsi C., Crocenzi T.S., Freire E., Roseman S.;

RT "Sugar transport by the bacterial phosphotransferase system.

RT Structural and thermodynamic domains of enzyme I of Salmonella

RT typhimurium.";

RL J. Biol. Chem. 266:19519-19527(1991).

RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN=LT2 / SCS1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

RA Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

RT Lt2.";

RL Nature 413:852-856(2001).

RN [3]

RN SEQUENCE OF 1-299 FROM N.A.

RC STRAIN=LT2;

RX MEDLINE=88257033; PubMed=3290198;

RA Byrne C.R., Monroe R.S., Ward K.A., Kredich N.M.;

RT "DNA sequences of the cysK regions of Salmonella typhimurium and

RT Escherichia coli and linkage of the cysK regions to ptsH.";

RL J. Bacteriol. 170:3150-3157(1988).

RN [4]

RN SEQUENCE OF 1-9 FROM N.A.

RC STRAIN=LT2;

RX MEDLINE=89237892; PubMed=2497295;

RA Schnierow B.J., Yamada M., Saier M.H. Jr.;

RT "Partial nucleotide sequence of the pts operon in Salmonella

RT typhimurium; comparative analyses in five bacterial genera.";

RL Mol. Microbiol. 3:113-118(1989).

CC -!- FUNCTION: THIS IS A COMPONENT OF THE PHOSPHOENOLPYRUVATE-DEPENDENT

CC SUGAR PHOSPHOTRANSFERASE SYSTEM (PTS), A MAJOR CARBOHYDRATE ACTIVE

CC -TRANSPORT SYSTEM. ENZYME I TRANSFERS THE PHOSPHORYL GROUP FROM

CC PHOSPHOENOLPYRUVATE (PEP) TO THE PHOSPHORYL CARRIER PROTEIN (HPR).

CC ENZYME I IS COMMON TO ALL PTS

CC -!- CATALYTIC ACTIVITY: Phosphoenolpyruvate + protein L-histidine -

CC Pyruvate + protein N(p1)-phospho-L-histidine.

CC -!- SUBUNIT: HOMODIMER.

CC -!- SIMILARITY: BELONGS TO THE PEP-UTILIZING ENZYMES FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M76176; AAA27060.1; -

DR

DR EMBL; AE008809; AAL21326.1; -

DR EMBL; M21450; AAA27053.1; -

DR EMBL; X14737; CAA32867.1; -

DR PIR; A41027; WQEBPI.

DR HSP; P08839; IEZC.

DR StyGene; SG10318; ptsI.

DR InterPro; IPR000121; PEP_utilizers.

DR Pfam; PF02896; PEP-utilizers.C; 1.

DR ProDom; PD000940; PEP_utilizers; 1.

DR PROSITE; PS00370; PEP_ENZYMES_PHOS_SITE; 1.

DR PROSITE; PS00742; PEP_ENZYMES_2; 1.

KW Phosphotransferase system; Transferase; Kinase; Sugar transport;

KW Phosphorylation; Complete proteome.

FT MOD_RES 189 189 PHOSPHORYLATION (BY SIMILARITY).

SQ SEQUENCE 575 AA; 63368 MW; 5A87EEE702D823F0 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 575;

Best Local Similarity 87.5%; Pred. No. 80;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLP 8

DB 193 MARSLELP 200

RESULT 21

VIVD_BPT7

ID VIVD_BPT7 STANDARD; PRT; 1318 AA.

AC P03726;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Internal virion protein D.

GN 16.

OS Bacteriophage T7.

OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;

OC T7-like phages.

OX NCBI_TaxID=10760;

RN [1]

RN SEQUENCE FROM N.A.

RX MEDLINE=83241725; PubMed=6864790;

RA Dunn J.J., Studier F.W.;

RT "Complete nucleotide sequence of bacteriophage T7 DNA and the

RT locations of T7 genetic elements.";

RL J. Mol. Biol. 166:477-535(1983).

RN [2]

RN SIMILARITY TO SLT.

RX MEDLINE=94262160; PubMed=8203016;

RA Koonin E.V., Rudd K.E.;

RT "A conserved domain in putative bacterial and bacteriophage

RT transglycosylases.";

RL Trends Biochem. Sci. 19:106-107(1994).

CC -!- FUNCTION: MAY BE INVOLVED IN THE LYSIS OF THE BACTERIAL CELL

CC WALL DURING THE RELEASE OF THE PHAGE PROGENY.

CC -!- SIMILARITY: BELONGS TO THE SLT FAMILY OF TRANSGLYCOSYLASES.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; V01146; CAA24434.1; -

DR PIR; A04352; HIBPD7.

DR PIR; S42332; S42332.

DR InterPro; IPR000189; SLT.

DR Pfam; PF01464; SLT; 1.

DR PROSITE; PS00922; TRANSGLYCOSYLASE; 1.

KW Cell wall; Hydrolase; Glycosidase.

FT DOMAIN 24 111 SLT-TYPE DOMAIN.

FT ACT_SITE 37 37 BY SIMILARITY.
SQ SEQUENCE 1318 AA; 143838 MW; 51A0AAA920CBF210 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 1318;

Best Local Similarity 75.0%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 1 MARSLLLP 8
||||:||
Db 1212 MARSTILP 1219

RESULT 22

ID 7LES_DROME STANDARD; PRT; 2554 AA.
AC P13368; OSU5V7; O9VZ36; Q9TY10;
DT 01-JAN-1990 (Rel. 13, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Sevenless protein (BC 2.7.1.112).
GN SEV OR HD-265 OR CG18085.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=88282538; PubMed=2840202;
RA Basler K., Hafen E.;
RT "Control of photoreceptor cell fate by the sevenless protein requires a functional tyrosine kinase domain.";
RL Cell 54:299-311 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=88329706; PubMed=3138161;
RA Bowtell D.L.L., Simon M.A., Rubin G.M.;
RT "Nucleotide sequence and structure of the sevenless gene of Drosophila melanogaster.";
RL Genes Dev. 2:620-634 (1988).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celiker S.E., Holt R.A., Evans C.A., Gocayne J.D., Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F., George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X., Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S., Bonkovsky D., Botchan M.R., Bouck J., Brokstein P., Brothier P., Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I., Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P., de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M., Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W., Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J., Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X., Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D., Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A., Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L., Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M., Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H., Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T., Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X., Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J., Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L., Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O., Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195 (2000).
RN [4]
RP SEQUENCE OF 2349-2408 FROM N.A.
RX MEDLINE=98401146; PubMed=9731193;
RA Oates A.C., Wollberg P., Achen M.G., Wilks A.F.;
RT "Sampling the genomic pool of protein tyrosine kinase genes using the polymerase chain reaction with genomic DNA.";
RL Biochem. Biophys. Res. Commun. 249:660-667 (1998).
RN [5]
RP IDENTIFICATION OF FN-III REPEATS.
RX MEDLINE=90199889; PubMed=2317871;
RA Norton P.A., Hynes R.O., Rens D.J.G.;
RT "Sevenless: haven found?";
RL Cell 61:15-16 (1990).
CC -1- FUNCTION: RECEPTOR FOR AN EXTRACELLULAR SIGNAL REQUIRED TO INSTRUCT A CELL TO DIFFERENTIATE INTO A R7 PHOTORECEPTOR. THE LIGAND FOR SEV IS THE BOSS (BRIDE OF SEVENLESS) PROTEIN ON THE SURFACE OF THE NEIGHBORING R8 CELL.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.
CC -1- SUBUNIT: MAY FORM A COMPLEX WITH DRK AND SOS.
CC -1- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-PROTEIN KINASES. SEVENLESS SUBFAMILY.
CC -1- SIMILARITY: CONTAINS 7 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC -1- CAUTION: UNCLEAR WHETHER THE POTENTIAL MEMBRANE SPANNING REGION NEAR THE N-TERMINUS IS PRESENT AS A TRANSMEMBRANE DOMAIN IN THE NATIVE PROTEIN OR SERVES AS A CLEAVED SIGNAL SEQUENCE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/) or send an email to license@isb-sib.ch).
CC -----
CC EMBL; J03158; AAA28882.1; -;
DR EMBL; X13666; CAA31960.1; ALT_INIT.
DR EMBL; X13666; CAB55310.1; -;
DR EMBL; AE003484; AAF47992.1; ALT_INIT.
DR EMBL; AJ002917; CAA05752.1; -;
DR PIR; A28912; TVPF7L.
DR HSP; P11362; IFGK.
DR FlyBase; FBgn0003366; sev.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR000033; Ldl_receptor_rep.
DR InterPro; IPR002011; Receptor_tyr_kin_II.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00041; fn3; 6.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00135; LY; 2.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; TyrcK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.
KW Transferrase; Tyrosine-protein kinase; Transmembrane; ATP-binding; Phosphorylation; Receptor; Vision; Repeat.


```
RT 1 2123 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 2124 2147 CYTOPLASMIC.
FT TRANSSEM 2148 2554 FIBRONECTIN TYPE-III 1.
FT DOMAIN 311 431 FIBRONECTIN TYPE-III 2.
FT DOMAIN 436 528 FIBRONECTIN TYPE-III 3.
FT DOMAIN 822 921 FIBRONECTIN TYPE-III 4.
FT DOMAIN 1298 1392 FIBRONECTIN TYPE-III 5.
FT DOMAIN 1680 1794 FIBRONECTIN TYPE-III 6.
FT DOMAIN 1797 1897 FIBRONECTIN TYPE-III 7.
FT DOMAIN 1898 1988 POLY-ARG.
FT DOMAIN 2038 2046 PROTEIN KINASE.
FT NP_BIND 2209 2485 ATP (BY SIMILARITY).
FT BINDING 2242 2242 ATP (BY SIMILARITY).
FT MOD_RES 2380 2380 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 30 30 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 129 129 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 481 481 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 505 505 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 617 617 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 647 647 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 966 966 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1228 1228 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1313 1313 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1353 1353 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1550 1550 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1557 1557 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1639 1639 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1725 1725 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1756 1756 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1804 1804 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1889 1889 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 1947 1947 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 2073 2073 N-LINKED (GLCNAC. .) (POTENTIAL).
FT MUTAGEN 2242 2242 K->M: INACTIVATES THE PROTEIN.
FT CONFLICT 392 392 V -> M (IN REF. 1).
FT CONFLICT 663 663 A -> T (IN REF. 3).
FT CONFLICT 1703 1703 N -> H (IN REF. 3).
FT CONFLICT 1730 1731 RG -> KE (IN REF. 3).
FT CONFLICT 1741 1741 V -> M (IN REF. 3).
FT CONFLICT 1823 1823 E -> Q (IN REF. 2).
FT CONFLICT 2271 2271 C -> R (IN REF. 1).
SQ SEQUENCE 2554 AA; 287022 MW; 09E238A0F27684F8 CRC64;

Query Match 73.2%; Score 30; DB 1; Length 2554;
Best Local Similarity 75.0%; Pred. No. 3.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLP 8
DB 772 LGRSLLLP 779

RESULT 23
IGF2_MUSVI
ID IGF2_MUSVI STANDARD; PRT; 129 AA.
AC P41694;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Insulin-like growth factor II precursor (IGF-II) (Fragment).
GN IGF2.
OS Mustela vison (American mink).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Mustelidae; Mustelinae;
OC Mustela.
OX NCBI_TaxID=9667;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=93307613; PubMed=768523;
RA Ekstroem T.J., Baecklin B.M., Lindqvist Y., Engstroem W.;
RT "Insulin-like growth factor II in the mink (Mustela vison):
```

```
RT determination of a cDNA nucleotide sequence and developmental
RT regulation of its expression."
RL Gen. Comp. Endocrinol. 90:243-250(1993).
CC -!- FUNCTION: THE INSULIN-LIKE GROWTH FACTORS POSSESS GROWTH-PROMOTING
CC ACTIVITY. IN VITRO, THEY ARE POTENT MITOGENS FOR CULTURED CELLS.
CC IGF-II IS INFLUENCED BY PLACENTAL LACTOGEN AND MAY PLAY A ROLE
CC IN FETAL DEVELOPMENT.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; S63459; AAB27392.2; -.
CC HSSP; P01344; IGF2.
CC InterPro: IPR000739; Insulin_IGF_relaxin.
CC Pfam; PF00049; Insulin; 1. Insulin_IGF_relaxin; 1.
CC ProDom; PD001048; Insulin; 1.
CC SMART; SM00078; IIGF; 1.
CC PROSITE; PS00262; INSULIN; 1.
CC Insulin family; Mitogen; Growth factor; Signal.
CC SIGNAL 1 24 BY SIMILARITY.
CC CHAIN 25 92 INSULIN-LIKE GROWTH FACTOR II.
CC DOMAIN 25 52 B.
CC DOMAIN 53 65 C.
CC DOMAIN 66 86 A.
CC DOMAIN 87 92 D.
CC PROPEP 93 >129 E PEPTIDE (BY SIMILARITY).
CC DISULFID 33 72 BY SIMILARITY.
CC DISULFID 45 85 BY SIMILARITY.
CC DISULFID 71 76 BY SIMILARITY.
CC NON_TER 129 129
SQ SEQUENCE 129 AA; 14436 MW; FD06661DAFB473D0 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 129;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
DB 5 MGKSLLLAPL 13

RESULT 24
OCRA_BACSU
ID OCRA_BACSU STANDARD; PRT; 167 AA.
AC P46911;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Menquinol-cytochrome C reductase iron-sulfur subunit (Rieske iron-
DE sulfur protein).
DE OCRA OR BFCa.
GN Bacillus subtilis.
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BR151;
RX MEDLINE=96074301; PubMed=7592464;
RA Yu J., Hederstedt L., Piggot P.J.;
RT "The cytochrome bc complex (menaquinone:cytochrome c reductase) in
RT Bacillus subtilis has a nontraditional subunit organization.";
RL J. Bacteriol. 177:6751-6760(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;
```

RA MEDLINE=96349105; PubMed=8760912;
RA Sorokin A.V., Azevedo V., Zumstein E., Galleron N., Ehrlich S.D.,
RA Serron P.;
RT "Sequence analysis of the Bacillus subtilis chromosome region between
RT the serA and kgd loci cloned in a yeast artificial chromosome.";
RL Microbiology 142:2005-2016(1996).
CC -!- FUNCTION: COMPONENT OF THE MENAQUINOL-CYTOCHROME C REDUCTASE
CC COMPLEX. THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S PROTEIN.
CC -!- SUBUNIT: THE MAIN SUBUNITS OF THE MENAQUINONE-CYTOCHROME C COMPLEX
CC ARE: CYTOCHROME B, THE RIESKE PROTEIN AND A CYTOCHROME B/C
CC SUBUNIT.
CC -!- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
CC BACTERIAL, CHLOROPLAST).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U25535; AAA85560.1; -;
DR EMBL: L47709; AAB38435.1; -;
DR EMBL: Z99115; CAB14172.1; -;
DR HSSP: P08980; IRFS.
DR Subtilist; BG11325; qcrA.
DR InterPro: IPR001281; Rieske.
DR Pfam: PF00355; Rieske.1.
DR PROSITE: PS00199; RIESKE_1; 1.
DR PROSITE: PS00200; RIESKE_2; 1.
KW Electron transport; Iron-sulfur; Oxidoreductase; Complete proteome.
FT METAL 100 100 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 102 102 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 121 121 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT METAL 124 124 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
FT DISULFID 105 123 BY SIMILARITY.
SQ SEQUENCE 167 AA; 18736 MW; 444F70B9BED2D143 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 167;
Best Local Similarity 55.6%; Pred. No. 38;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLPL 9
Db 24 MAASMLPM 32
||| | | | | |

RESULT 25
NIRQ_PSEAE STANDARD; PRT; 260 AA.
AC Q51481;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Denitrification (Rel. 40, Last annotation update)
GN NIRQ OR PA0520.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
CC Pseudomonas.
OX NCBI_TaxID=287;
[1]
SEQUENCE FROM N.A.
RP STRAIN=PA01161;
RX MEDLINE=94362287; PubMed=7765251;
RA Arai H., Igarashi Y., Kodama T.;
RT "Structure and ANR-dependent transcription of the nir genes for
RT denitrification from Pseudomonas aeruginosa.";
RL Biosci. Biotechnol. Biochem. 58:1286-1291(1994).
[2]
SEQUENCE FROM N.A.
RP STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;

RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltz L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Speller D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen";
RL Nature 406:959-964(2000).
CC -!- FUNCTION: ACTIVATOR OF NITRITE AND NITRIC OXIDE REDUCTASES.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- INDUCTION: UNDER DENITRIFYING CONDITIONS.
CC -!- SIMILARITY: BELONGS TO THE CBQ/NIRQ/NORQ/GPNV FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: D37883; BAA07123.1; -;
DR EMBL: AE004489; AAG03909.1; -;
KW ATP-binding; DNA-binding; Transcription regulation; Activator;
FT Complete proteome.
FT NP_BIND 32 39 ATP (POTENTIAL).
FT NP_BIND 92 99 ATP (POTENTIAL).
FT DNA_BIND 234 253 H-T-H MOTIF (BY SIMILARITY).
SQ SEQUENCE 260 AA; 28904 MW; 3FD36F19BEB438B5 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 260;
Best Local Similarity 77.8%; Pred. No. 59;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLPL 9
Db 45 MARRLELP 53
||| | | | | |

RESULT 26
PYGL_SYNEL STANDARD; PRT; 276 AA.
AC P50039;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phycobilisome rod-core linker polypeptide cpcgl.
GN CPCGL.
OS Synechococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.
OX NCBI_TaxID=32046;
[1]
SEQUENCE FROM N.A.
RP Shimazu T., Soga M., Hirano M., Katoh S.;
RT "Cloning and sequencing of the phycocyanin operon from the
RT thermophilic cyanobacterium Synechococcus elongatus";
RL Submitted (SEP-1992) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ROD-CORE LINKER PROTEIN REQUIRED FOR ATTACHMENT OF
CC PHYCOCYANIN TO ALLOPHYCOCYANIN IN CORES OF PHYCABILISOMES.
CC -!- FUNCTION: LINKER POLYPEPTIDES DETERMINE THE STATE OF AGGREGATION
CC AND THE LOCATION OF THE DISC-SHAPED PHYCABILIPROTEIN UNITS WITHIN
CC THE PHYCABILISOME AND MODULATE THEIR SPECTROSCOPIC PROPERTIES IN
CC ORDER TO MEDIATE A DIRECTED AND OPTIMAL ENERGY TRANSFER.
CC -!- SUBUNIT: THE PHYCABILISOME IS A HEMIDISCAL STRUCTURE THAT IS
CC COMPOSED OF TWO DISTINCT SUBSTRUCTURES: A CORE COMPLEX (THAT
CC CONTAINS PHYCABILIPROTEINS) AND A NUMBER OF RODS RADIATING FROM
CC THE CORE.
CC -!- SIMILARITY: TO OTHER PHYCABILISOME LINKER PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; D13173; BAA02461.1; -;
 DR InterPro: IPR001297; PBS_linker_poly.
 DR Pfam; PF00427; PBS_linker_poly; 1.
 KW Phycobilisome; Photosynthesis; Multigene family.
 FT INIT_MET 0 BY SIMILARITY.
 SQ SEQUENCE 276 AA; 31493 MW; 56D4EEC582AA762E CRC64;

Query Match 70.7%; Score 29; DB 1; Length 276;
 Best Local Similarity 62.5%; Pred. No. 63;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLP 8
 DB 219 MARSLVP 226
 {}||||:|

RESULT 27

TH12_SCHPO STANDARD; PRT; 328 AA.
 AC P40998; O74785;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Thiazole biosynthetic enzyme, mitochondrial precursor.
 GN TH12 OR NMT2 OR SPC36H8.01.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=95084634; PubMed=7992507;
 RA Manetti A.G.O., Rosetto M., Maundrell K.G.;
 RT "nmt2 of fission yeast: a second thiamine-repressible gene co-
 RT ordinately regulated with nmt1.";
 RL Yeast 10:1075-1082(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Wood V., Rajandream M.A., Barrell B.G., Devlin K., Churcher C.M.;
 RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLY INVOLVED IN THE BIOSYNTHESIS OF THE THIAZOLE
 CC MOIETY OF THE THIAMINE MOLECULE.
 CC -!- SUBCELLULAR LOCATION: Mitochondrial (Potential).
 CC -!- INDUCTION: REPRESSED BY THIAMINE.
 CC -!- SIMILARITY: BELONGS TO THE TH14 FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; X82363; CAA57779.1; -;
 DR EMBL; AL031743; CAA21093.1; -;
 DR PIR; S45597; S45597.
 DR InterPro: IPR001327; FAD_pyf_redox.
 DR InterPro: IPR002922; Th14.
 DR Pfam; PF01946; Th14; 1.
 DR PRINTS; PR00368; FADPNR.
 DR Thiamine biosynthesis; Mitochondrion; Transist peptide; FAD; NAD.
 KW TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
 FT CHAIN ? 328 THIAZOLE BIOSYNTHETIC ENZYME.
 FT NP_BIND 78 108 FAD OR NAD (POTENTIAL).

FT CONFLICT 299 328 FGMFMFGIKAAQEAIAIFDERKAVNERYL -> SVV (IN
 FT REF. 1).
 SQ SEQUENCE 328 AA; 35274 MW; BF18CCB367BEC421 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 328;
 Best Local Similarity 75.0%; Pred. No. 75;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLP 8
 DB 164 MARTIALP 171
 ||||:|

RESULT 28

CLCB_PSEPU STANDARD; PRT; 370 AA.
 ID CLCB_PSEPU
 AC P11452; P15741;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Chloromuconate cycloisomerase (EC 5.5.1.7) (Muconate cycloisomerase
 DE II).
 GN CLCB.
 OS Pseudomonas putida.
 OG Plasmid pAC27.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AC867;
 RX MEDLINE=90060834; PubMed=2583528;
 RA Ghosal D., You I.-S.;
 RT "Operon structure and nucleotide homology of the chlorocatechol
 RT oxidation genes of plasmids pJP4 and pAC27.";
 RL Gene 83:225-232(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87260828; PubMed=3299368;
 RA Prantz B., Chakrabarty A.M.;
 RT "Organization and nucleotide sequence determination of a gene cluster
 RT involved in 3-chlorocatechol degradation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:4460-4464(1987).
 CC -!- CATALYTIC ACTIVITY: 2-CHLORO-2,5-DIHYDRO-5-OXOFURAN-2-ACETATE -
 CC 3-CHLORO-CIS-CIS-MUCONATE (SPONTANEOUS ELIMINATION OF HCL PRODUCES
 CC CIS-4-CARBOXYMETHYLENEBUT-2-EN-4-OLIDE) (ALSO REACTS IN REVERSE
 CC DIRECTION ON 2-CHLORO-CIS,CIS-MUCONATE).
 CC -!- PATHWAY: 3-CHLOROCATECHOL DEGRADATION.
 CC -!- MISCELLANEOUS: CHLOROMUCONATE CYCLOISOMERASE II IS HIGHLY ACTIVE
 CC TOWARD CHLORINATED SUBSTRATES BUT RETAINS DIMINISHED ACTIVITY
 CC TOWARD THE NONCHLORINATED SUBSTRATES.
 CC -!- SIMILARITY: BELONGS TO THE MANDELATE RACEMASE / MUCONATE
 CC LACTONIZING ENZYME FAMILY.

CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; M31457; AAA98260.1; -;
 DR EMBL; M16964; AAA98282.1; -;
 DR PIR; B27316; B27316.
 DR PIR; B27058; B27058.
 DR PIR; JQ0176; JQ0176.
 DR HSP; P05404; 2CHR.
 DR InterPro: IPR001354; MR_MLE.
 DR Pfam; PF01188; MR_MLE; 1.
 DR Pfam; PF02746; MR_MLE_N; 1.
 DR PROSITE; PS00908; MR_MLE_1; 1.
 DR PROSITE; PS00909; MR_MLE_2; 1.

KW Aromatic hydrocarbons catabolism: Isomerase; Plasmid.
FT CONFLICT 42 V -> C (IN REF. 2).
FT CONFLICT 114 R -> H (IN REF. 2).
SQ SEQUENCE 370 AA; 39846 MW; 9279DD393F42BBCD CRC64;

Query Match 70.7%; Score 29; DB 1; Length 370;
Best Local Similarity 87.5%; Pred. No. 84;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9

Db 113 ARSLNPL 120

RESULT 29

HSLU_THEME STANDARD; PRT; 463 AA.

AC QWY22;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE ATP-dependent hsl protease ATP-binding subunit hslu.
GN HSLU OR TM0522;
OS Thermotoga maritima.
OC Bacteria; Thermotogales; Thermotoga.
OX NCBI_TaxID=2336;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSB8 / DSM 3109;
RX MEDLINE=99287316; PubMed=10360571;
RA Nelson K.E., Clayton R.A., Gill S.R., Gwinn M.L., Dodson R.J.,
Haft D.H., Hickey E.K., Peterson J.D., Nelson W.C., Ketchum K.A.,
McDonald L., Utterback T.R., Malek J.A., Linher K.D., Garrett M.M.,
Stewart A.M., Cotton M.D., Pratt M.S., Phillips C.A., Richardson D.,
Heidelberg J., Sutton G.G., Fleischmann R.D., Eisen J.A., White O.,
Saizberg S.B., Smith H.O., Venter J.C., Fraser C.M.;
RT "Evidence for lateral gene transfer between Archaea and Bacteria from genome sequence of Thermotoga maritima.";
RL Nature 399:323-329(1999).

CC -!- FUNCTION: CHAPERONE SUBUNIT OF A PROTEASOME-LIKE DEGRADATION

CC COMPLEX (BY SIMILARITY).

CC -!- SUBUNIT: INTERACTS WITH HSLV (BY SIMILARITY).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).

CC -!- SIMILARITY: BELONGS TO THE CLPX CHAPERONE FAMILY. HSLU SUBFAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; A5001728; AAD35607.1; -.

DR HSSP; P32166; ID02.

DR TIGR; TM0522; -.

DR InterPro; IPR003593; AAA.

DR InterPro; IPR003959; AAA_subfam.

DR Pfam; PF00004; AAA.1.

DR SMART; SM00382; AAA.1.

KW Chaperone; ATP-binding; Complete proteome.

FT NP_BIND 60 67 ATP (POTENTIAL).

SQ SEQUENCE 463 AA; 53052 MW; F871CD909FCBA5CA CRC64;

Query Match 70.7%; Score 29; DB 1; Length 463;

Best Local Similarity 75.0%; Pred. No. 11e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9

Db 241 ARKVLPL 248

RESULT 30

ID INUL_KLUMA

AC P28999; STANDARD; PRT; 555 AA.

DT 01-DEC-1992 (Rel. 24, Created)

DT 01-DEC-1992 (Rel. 24, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Inulinase precursor (EC 3.2.1.7) (2,1-beta-D-fructanfructanohydrolase)

DE (Inulase).

GN INUL.

OS Kluyveromyces marxianus (Yeast) (Kluyveromyces fragilis).

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.

OX NCBI_TaxID=4911;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE OF 24-56.

RC STRAIN=ATCC 12424;

RX MEDLINE=91372407; PubMed=1840529;

RA Laloux O., Cassart J.-P., van Beeumen J., Delcour J., Vandenhoute J.;

RT "Cloning and sequencing of the inulinase gene of Kluyveromyces

RT marxianus var. marxianus ATCC 12424.";

RL FEBS Lett. 289:64-68(1991).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=CBS 6556;

RX MEDLINE=94128347; PubMed=7764385;

RA Bergkamp R.J.M., Bootsman T.C., Toschka H.Y., Moeren A.T.A., Kox L.,

Verbakel J.M.A., Geerse R.H., Planta R.J.;

RT "Expression of an alpha-galactosidase gene under control of the

RT homologue inulinase promoter in Kluyveromyces marxianus.";

RL Appl. Microbiol. Biotechnol. 40:309-317(1993).

RN [3]

RP SEQUENCE OF 24-43

RX MEDLINE=92304047; PubMed=2135869;

RA Rouwenhorst R.J., Hensing M., Verbakel J., Scheffers W.A.,

van Dijken J.P.;

RT "Structure and properties of the extracellular inulinase of

RT Kluyveromyces marxianus CBS 6556.";

RL Appl. Environ. Microbiol. 56:3337-3345(1990).

CC -!- FUNCTION: HAS BOTH INULASE AND INVERTASE ACTIVITY.

CC -!- CATALYTIC ACTIVITY: Endohydrolysis of 2,1-beta-D-fructosidic

CC linkages in inulin.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: BELONGS TO FAMILY 32 OF GLYCOSYL HYDROLASES.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; X57202; CAA0488.1; -.

DR EMBL; X68479; CAA48500.1; -.

DR PIR; S17502; S17502.

DR InterPro; IPR001362; Glyco_hydro_32.

DR Pfam; PF00251; Glyco_hydro_32.1.

DR PROSITE; PS00609; GLYCOSYL_HYDROL_F32.1.

KW Hydrolase; Glycosidase; Glycoprotein; signal.

FT SIGNAL 1 15 OR 16, OR 17 (POTENTIAL).

FT PROPEP 16 23 POTENTIAL.

FT CHAIN 24 555 INULINASE.

FT ACT_SITE 53 53 BY SIMILARITY.

FT CARBOHYD 32 32 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 122 122 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 123 123 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 129 129 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 176 176 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 226 226 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 255 255 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).

```
FT CARBOHYD 385 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 393 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 490 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 549 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 3 F -> L (IN STRAIN CBS 6556).
FT VARIANT 165 E -> Q (IN STRAIN CBS 6556).
FT VARIANT 168 T -> S (IN STRAIN CBS 6556).
FT VARIANT 251 V -> A (IN STRAIN CBS 6556).
FT VARIANT 252 D -> DS (IN STRAIN CBS 6556).
FT VARIANT 299 G -> D (IN STRAIN CBS 6556).
FT VARIANT 479 T -> N (IN STRAIN CBS 6556).
FT CONFLICT 24 D -> S (IN REF. 3).
FT CONFLICT 43 H -> Y (IN REF. 3).
SQ SEQUENCE 555 AA; 62213 MW; 723BAADDCC3BF0907 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 555;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
DB 4 AYSLLPL 11

RESULT 31
MMLD_STRCO STANDARD; PRT; 705 AA.
AC Q9XA86;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative membrane protein SCF43A.29C.
GN SCF43A.29C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Seeger K., Harris D., James K.D., Parkhill J., Barrell B.G.,
RA Rajandream M.A.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE MMLP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AL096837; CAB48916.1;
CC DR InterPro: IPR000731; HMGCR_patched_5TM.
CC DR PROSITE: PS50156; SSD; 2.
CC KW Hypothetical protein; Transmembrane.
FT TRANSMEM 16 36 POTENTIAL.
FT TRANSMEM 177 197 POTENTIAL.
FT TRANSMEM 202 222 POTENTIAL.
FT TRANSMEM 237 257 POTENTIAL.
FT TRANSMEM 281 301 POTENTIAL.
FT TRANSMEM 316 336 POTENTIAL.
FT TRANSMEM 373 393 POTENTIAL.
FT TRANSMEM 529 549 POTENTIAL.
FT TRANSMEM 554 574 POTENTIAL.
FT TRANSMEM 587 607 POTENTIAL.
FT TRANSMEM 627 647 POTENTIAL.
FT TRANSMEM 648 668 POTENTIAL.
SQ SEQUENCE 705 AA; 73316 MW; DF4C3B77AE2EA2C4 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 705;
Best Local Similarity 87.5%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
```

```
Best Local Similarity 85.7%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLLPL 9
DB 199 RSVLLPL 205

RESULT 32
CATA_HALMA STANDARD; PRT; 731 AA.
ID CATA_HALMA
AC O59651;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Peroxidase/catalase (EC 1.11.1.6) (Catalase-peroxidase).
GN PERA.
OS Haloarcula marismortui (Halo bacterium marismortui).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae; Haloarcula.
OX NCBI_TaxID=2238;
RN [1]
RP SEQUENCE FROM N.A.
RA Franzetti B., Cannac-Caffrey V., Petillot Y., Hudry-Clergeon G.,
RA Gagnon J., Zaccal G.;
RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BIFUNCTIONAL, EXHIBITING BOTH A CATALASE AND
CC BROAD-SPECTRUM PEROXIDASE ACTIVITIES (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: DONOR + H(2)O(2) -> OXIDIZED DONOR + 2 H(2)O.
CC -1- COFACTOR: TWO PROTOHEME IX GROUPS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE PEROXIDASE FAMILY. BACTERIAL
CC PEROXIDASE/CATALASE SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Y16851; CAA76423.1;
CC DR HSSP; P00431; ICCK.
CC DR InterPro: IPR002016; Peroxidase.
CC DR Pfam: PF00141; peroxidase; 1.
CC DR PRINTS: PR00458; PEROXIDASE
CC DR PROSITE: PS00435; PEROXIDASE_1; 1.
CC DR PROSITE: PS00436; PEROXIDASE_2; 1.
CC KW Oxidoreductase; Peroxidase; Iron; Heme; Hydrogen peroxide.
FT ACT_SITE 92 92 BY SIMILARITY.
FT ACT_SITE 96 96 BY SIMILARITY.
FT BINDING 259 259 PROXIMAL HEME LIGAND (BY SIMILARITY).
SQ SEQUENCE 731 AA; 81383 MW; 0E12DE0CF72FFA3 CRC64;

Query Match 70.7%; Score 29; DB 1; Length 731;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
DB 132 ARLLPLI 139

RESULT 33
PHS2_DICDI STANDARD; PRT; 992 AA.
ID PHS2_DICDI
AC P34114;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1994 (Rel. 30, Last annotation update)
DE Glycogen phosphorylase 2 (EC 2.4.1.1) (GP2).
GN GLPD.
OS Dictyostellium discoideum (Slime mold).
OS Eukaryota; Mycetozoa; Dictyostelida; Dictyostellium.
```

```

OX NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX STRAIN=AX3;
RC MEDLINE=92129303; PubMed=1310312;
RA Rutherford C.L., Peery R.B., Sucic J.F., Yin Y., Rogers P.V.,
RA Luo S., Selmin O.;
RT "Cloning, structural analysis, and expression of the glycogen
RT phosphorylase-2 gene in Dictyostelium.";
RL J. Biol. Chem. 267:2294-2302(1992).
CC -1- FUNCTION: PHOSPHORYLASE IS AN IMPORTANT ALLOSTERIC ENZYME IN
CC CARBOHYDRATE METABOLISM. ENZYMES FROM DIFFERENT SOURCES DIFFER IN
CC THEIR REGULATORY MECHANISMS AND IN THEIR NATURAL SUBSTRATES.
CC HOWEVER, ALL KNOWN PHOSPHORYLASES SHARE CATALYTIC AND STRUCTURAL
CC PROPERTIES.
CC -1- CATALYTIC ACTIVITY: [(1,4)-alpha-D-glucosyl](N) + phosphate =
CC [(1,4)-alpha-D-glucosyl](N-1) + alpha-D-glucose 1-phosphate.
CC -1- COFACTOR: PYRIDOXAL PHOSPHATE.
CC -1- SUBUNIT: HOMODIMER.
CC -1- DEVELOPMENTAL STAGE: APPEARS DURING CELL DIFFERENTIATION; ABSENT
CC IN AMOEBAE AND EARLY STAGES OF DEVELOPMENT, REACHES A MAXIMUM
CC LEVEL OF EXPRESSION AT THE SLUG STAGE AND THEN DECREASES.
CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC -1- PTM: ENZYME ACTIVITY REQUIRES PROCESSING OF THE 113 kDa PEPTIDE TO
CC AN ENZYMICALLY ACTIVE 106 kDa FORM OF THE PROTEIN. PROCESSING
CC WOULD OCCUR NEAR THE MIDDLE OF THE GLN-RICH REPETITIVE ELEMENT.
CC -1- MISCELLANEOUS: IN D.DISCOIDEUM GLYCOGEN PHOSPHORYLASE EXISTS AS 2
CC DEVELOPMENTALLY REGULATED FORMS, GP1 AND GP2, WHICH ARE THE
CC PRODUCTS OF SEPARATE GENES.
CC -1- SIMILARITY: BELONGS TO THE GLYCOGEN PHOSPHORYLASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M77492; AAA33211.1; -
CC PIR; A42318; A42318.
CC HSP; P06738; 1YGP.
CC Dictyob; DD03017; glpD.
CC InterPro; IPR000811; Phosphorylase.
CC Pfam; PF00343; phosphorylase; 1.
CC PROSITE; PS00102; PHOSPHORYLASE; 1.
CC Transferrase; Glycosyltransferase; Carbohydrate metabolism;
CC Glycogen metabolism; Allosteric enzyme; Pyridoxal phosphate;
CC Multigene family.
CC DOMAIN 44 75 GLN-RICH.
CC SITE 229 229 INVOLVED IN THE ASSOCIATION OF SUBUNITS
CC (BY SIMILARITY).
CC ACT_SITE 242 242 MAY BE INVOLVED IN THE ALLOSTERIC CONTROL
CC OF ENZYME ACTIVITY (BY SIMILARITY).
CC BINDING 762 762 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC SEQUENCE 992 AA; 112542 MW; 27E86AC6783FAF72 CRC64;
CC
CC Query Match 70.7% Score 29; DB 1; Length 992;
CC Best local Similarity 85.7% Pred. No. 2.2e+02;
CC Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC
CC Qy 2 ARSLLLP 8
CC I I I I I I
CC Db 922 ARSLLLVP 928
CC
CC RESULT 34
CC NIDO_HUMAN
CC ID NIDO_HUMAN STANDARD; PRT; 1247 AA.
CC AC P14543; Q14942;
CC DT 01-JAN-1990 (Rel. 13, Created)
CC DT 01-JAN-1990 (Rel. 13, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE Nidogen precursor (Entactin).
GN NID.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90091745; PubMed=2574658;
RA Nagayoshi T., Sanborn D., Hickok N.J., Olsen D.R., Fazio M.J.,
RA Chu M.-L., Knowlton R., Mann K., Deutzmann R., Timpl R., Uitto J.;
RT "Human nidogen: complete amino acid sequence and structural domains
RT deduced from cDNAs, and evidence for polymorphism of the gene.";
RL DNA 8:581-594(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96044428; PubMed=7557988;
RA Zimmermann K., Hoischen S., Hafner M., Nischt R.;
RT "Genomic sequences and structural organization of the human nidogen
RT gene (NID).";
RL Genomics 27:245-250(1995).
RN [3]
RP SEQUENCE OF 667-1247 FROM N.A.
RC TISSUE=Placenta;
RX MEDLINE=89270475; PubMed=2471408;
RA Olsen D.R., Nagayoshi T., Fazio M., Mattei M.-G., Weiss D.,
RA Timpl R., Chu M.-L., Uitto J.;
RT "Human nidogen: cDNA cloning, cellular expression, and mapping of the
RT gene to chromosome 1q43.";
RL Am. J. Hum. Genet. 44:876-885(1989).
CC -1- FUNCTION: SULFATED GLYCOPROTEIN WHICH IS WIDELY DISTRIBUTED IN
CC BASEMENT MEMBRANES AND THAT IS TIGHTLY ASSOCIATED WITH LAMININ.
CC ALSO BINDS TO COLLAGEN IV. IT PROBABLY HAS A ROLE IN CELL-
CC EXTRACELLULAR MATRIX INTERACTIONS.
CC -1- SUBCELLULAR LOCATION: BASEMENT MEMBRANES.
CC -1- PTM: N- AND O-GLYCOSYLATED.
CC -1- SIMILARITY: CONTAINS 1 THYROGLOBULIN TYPE-I DOMAIN.
CC -1- SIMILARITY: CONTAINS 5 LDL-RECEPTOR YWTD DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M30269; AAA59932.1; -
CC EMBL; X82245; CAA57709.1; -
CC EMBL; X84819; CAA57709.1; JOINED.
CC EMBL; X84820; CAA57709.1; JOINED.
CC EMBL; X84821; CAA57709.1; JOINED.
CC EMBL; X84822; CAA57709.1; JOINED.
CC EMBL; X84823; CAA57709.1; JOINED.
CC EMBL; X84824; CAA57709.1; JOINED.
CC EMBL; X84825; CAA57709.1; JOINED.
CC EMBL; X84826; CAA57709.1; JOINED.
CC EMBL; X84827; CAA57709.1; JOINED.
CC EMBL; X84828; CAA57709.1; JOINED.
CC EMBL; X84829; CAA57709.1; JOINED.
CC EMBL; X84830; CAA57709.1; JOINED.
CC EMBL; X84831; CAA57709.1; JOINED.
CC EMBL; X84832; CAA57709.1; JOINED.
CC EMBL; X84833; CAA57709.1; JOINED.
CC EMBL; X84834; CAA57709.1; JOINED.
CC EMBL; X84835; CAA57709.1; JOINED.
CC EMBL; X84836; CAA57709.1; JOINED.
CC EMBL; X84837; CAA57709.1; JOINED.
CC EMBL; M27445; AAA57261.1; -
CC PIR; A33322; MHUND.
CC HSP; P07204; IADX.
CC MIM; 131390; -

```

DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR001881; EGF Ca.
DR InterPro: IPR000033; Ldl_receptor_rep.
DR InterPro: IPR003886; Nidogen_ext.
DR InterPro: IPR000716; Thyroglobulin_1.
DR Pfam: PF00008; EGF; 6.
DR Pfam: PF00058; ldl_recept_b; 3.
DR Pfam: PF00086; thyroglobulin_1; 1.
DR SMART: SM00179; EGF_CA; 2.
DR SMART: SM00001; EGF-like; 4.
DR SMART: SM00135; LY; 5.
DR SMART: SM00539; NIDO; 1.
DR SMART: SM00211; TY; 1.
DR PROSITE: PS00010; ASX_HYDROXYL; 3.
DR PROSITE: PS00484; THYROGLOBULIN_1; 1.
DR PROSITE: PS00022; EGF_1; 1.
DR PROSITE: PS01186; EGF_2; 5.
DR PROSITE: PS01187; EGF_CA; 2.
KW Basement membrane; Extracellular matrix; Glycoprotein; Sulfation;
Signal; Calcium-binding; Repeat; EGF-like domain; Cell adhesion.
FT SIGNAL 1 28
FT CHAIN 29 1247
FT DOMAIN 29 669
FT DOMAIN 670 917
FT DOMAIN 918 1247
FT DOMAIN 386 426
FT DOMAIN 668 709
FT DOMAIN 710 751
FT DOMAIN 758 801
FT DOMAIN 802 840
FT DOMAIN 872 919
FT DOMAIN 989 1030
FT DOMAIN 1032 1073
FT DOMAIN 1075 1118
FT DOMAIN 1124 1163
FT DOMAIN 1208 1244
FT MOD_RES 289 289
FT MOD_RES 296 296
FT DISULFID 672 685
FT DISULFID 679 695
FT DISULFID 697 708
FT DISULFID 714 727
FT DISULFID 721 736
FT DISULFID 738 750
FT DISULFID 762 777
FT DISULFID 769 787
FT DISULFID 789 800
FT DISULFID 806 817
FT DISULFID 811 826
FT DISULFID 828 839
FT DISULFID 1212 1223
FT DISULFID 1219 1232
FT DISULFID 1234 1243
FT SITE 702 704
FT CARBOHYD 1137 1137
FT CONFLICT 33 34
FT CONFLICT 37 42
FT CONFLICT 1115 1115
SQ SEQUENCE 1247 AA; 136488 MW; 4681B5B3CEC1575B CRC64;
Query Match 70.7%; Score 29; DB 1; Length 1247;
Best Local Similarity 85.7%; Pred. No. 2.8e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 3 RSLLLPL 9
Db 14 RALLPL 20
RESULT 35
EXPA_DROME STANDARD; PRT; 1429 AA.
ID

AC Q07436;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Expanded protein.
GN EX.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephygoidae; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX TISSUE=Imaginal disks;
RC MEDLINE=94094747; PubMed=8269855;
RA Boedigheimer M., Laughon A.;
RT "Expanded: a gene involved in the control of cell proliferation in
imaginal discs";
RL Development 118:1291-1301(1993).
RN [2]
RP REVISIONS.
RA Boedigheimer M.;
RL Submitted (DEC-1996) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: INVOLVED IN THE CONTROL OF CELL PROLIFERATION IN
IMAGINAL DISCS. MAY BIND TO CERTAIN PROTEINS OF SIGNAL
TRANSDUCTION PATHWAYS BY INTERACTION WITH THEIR SH3 DOMAINS.
CC -!- SUBCELLULAR LOCATION: APICAL SURFACE OF DISC CELLS.
CC -!- DISEASE: MUTATIONS OF EXPANDED PROTEIN CAUSE HYPERPLASIA OF THE
IMAGINAL DISC RESULTING IN WING OVERGROWTH. THIS OVERGROWTH IS
LIMITED TO SPECIFIC REGIONS ALONG THE 2 WING AXES. DEFECTS ALSO
IN EYES, HEAD, THORAX AND LIMBS WHERE DUPLICATION AND BULGING
OFTEN OCCUR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L14768; AAB39774.1; -.
DR FlyBase; FBgn0004583; ex.
DR InterPro; IPR000299; Band_4.1.
DR Pfam; PF00373; Band_4.1; 1.
DR SMART; SM00295; B41_1.
DR PROSITE; PS50057; BAND_4.1_3; UNKNOWN_1.
KW Developmental protein; SH3-binding.
FT SITE 1008 1016 SH3-BINDING (POTENTIAL).
FT SITE 1012 1020 SH3-BINDING (POTENTIAL).
FT SITE 1149 1157 SH3-BINDING (POTENTIAL).
FT DOMAIN 409 412 POLY-GLU.
FT DOMAIN 782 788 POLY-PRO.
FT DOMAIN 952 955 POLY-HIS.
FT DOMAIN 1002 1005 POLY-PRO.
FT DOMAIN 1011 1017 POLY-PRO.
FT DOMAIN 1081 1084 POLY-PRO.
FT DOMAIN 1149 1154 POLY-PRO.
FT DOMAIN 1158 1168 POLY-ALA.
FT DOMAIN 1170 1174 POLY-SER.
FT DOMAIN 1199 1205 POLY-SER.
FT DOMAIN 1416 1424 POLY-GLN.
SQ SEQUENCE 1429 AA; 153886 MW; 3CB08D2FC4862062 CRC64;
Query Match 70.7%; Score 29; DB 1; Length 1429;
Best Local Similarity 85.7%; Pred. No. 3.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 ARSLLLP 8
Db 426 SRSLLLP 432


```

SQ SEQUENCE 192 AA; 21095 MW; ED8E9B51AE4909BE CRC64;
  Query Match 68.3%; Score 28; DB 1; Length 192;
  Best Local Similarity 75.0%; Pred.No. 71;
  Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
   ||:|||||
Db 26 ARNYLLPL 33

RESULT 39
TEF_PHOSU
ID TEF_PHOSU STANDARD; PRT; 221 AA.
AC P97516;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE 30-MAY-2000 (Rel. 39, Last annotation update)
GN Thymotroph embryonic factor (Fragment).
OS Phodopus (Striped hairy-footed hamster) (Djungarian hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Cricetinae;
OC Phodopus.
OX NCBI_TaxID=10044;
RN [1]
RP SEQUENCE FROM N.A.
RA Bockmann J.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO AND TRANSACTIVATES
CC THE TSHB PROMOTER. BINDS TO A MINIMAL DNA-BINDING SEQUENCE 5'-
CC [TC]AG[AG]TAA[TC][AG]-3' (BY SIMILARITY).
CC -!- SUBUNIT: BINDS DNA AS A HOMODIMER OR A HETERODIMER. CAN FORM A
CC HETERODIMER WITH DBP (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC -!- SIMILARITY: BELONGS TO THE BZIP FAMILY. PAR SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Y11149; CAA72036.1;
CC InterPro: IPR001871; BZIP.
CC SMART; SM00338; BRLZ; 1.
KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
FT NON_TER 1
FT DOMAIN 106 157 PRO-RICH (PROLINE/ACIDIC REGION (PAR)).
FT DNA_BIND 179 197 BASIC MOTIF
FT DOMAIN 208 >221 LEUCINE-ZIPPER.
FT NON_TER 221
SQ SEQUENCE 221 AA; 24528 MW; 69645BCA042CFA8B CRC64;
  Query Match 68.3%; Score 28; DB 1; Length 221;
  Best Local Similarity 55.6%; Pred.No. 82;
  Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
   ||:|||||
Db 57 LAQNLLPLV 65

RESULT 40
SSAT_SALTY
ID SSAT_SALTY STANDARD; PRT; 259 AA.
AC P96068;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Secretion system apparatus protein ssat.

GN SSAT OR STM1421.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=97285756; PubMed=9140973;
RA Hensel M., Shea J.E., Raupach B., Monack D., Falkow S., Gleeson C.,
RA Kubo T., Holden D.W.;
RT "Functional analysis of ssaj and the ssak/U operon, 13 genes encoding
RT components of the type III secretion apparatus of Salmonella
RT Pathogenicity Island 2.";
RL Mol. Microbiol. 24:155-167(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
RX MEDLINE=21534948; PubMed=11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2.";
RL Nature 413:852-856(2001).
CC -!- FUNCTION: PART OF A TYPE III SECRETION SYSTEM.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE FLIR/MOPE/SPAR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X99944; CAA68201.1;
CC EMBL: AE008761; AAL20345.1;
CC StyGene; SGI0717; ssat.
CC InterPro: IPR002010; Bac_export_1.
CC Pfam; PF01311; Bac_export_1; 1.
CC PRINTS; PR00953; TYPE3IMRPROT.
KW Transport; Protein transport; Transmembrane; Complete proteome.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 35 55 POTENTIAL.
FT TRANSMEM 78 98 POTENTIAL.
FT TRANSMEM 127 147 POTENTIAL.
FT TRANSMEM 185 205 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
SQ SEQUENCE 259 AA; 29218 MW; 8E5F0734991373DD CRC64;
  Query Match 68.3%; Score 28; DB 1; Length 259;
  Best Local Similarity 66.7%; Pred.No. 96;
  Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
   ||:|||||
Db 20 LSLSLLLPL 28

RESULT 41
TEF_RAT
ID TEF_RAT STANDARD; PRT; 261 AA.
AC P41224;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Thymotroph embryonic factor.
GN TEF.
OS Rattus norvegicus (Rat).
```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=pituitary; PubMed=1916262;
RX MEDLINE=92009166;
RA Drolet D.W., Scully K.M., Simmons D.M., Wegner M., Chu K.,
RA Swanson L.W., Rosenfeld M.G.;
RT "TGF, a transcription factor expressed specifically in the anterior
RT pituitary during embryogenesis, defines a new class of leucine zipper
RT proteins";
RL Genes Dev. 5:1739-1753(1991).
CC -!- FUNCTION: TRANSCRIPTION FACTOR THAT BINDS TO AND TRANSACTIVATES
CC THE TSHB PROMOTER. BINDS TO A MINIMAL DNA-BINDING SEQUENCE 5'-
CC [TC][AG][AG]TTA[TC][AG]-3'.
CC -!- SUBUNIT: BINDS DNA AS A HOMODIMER OR A HETERODIMER. CAN FORM A
CC HETERODIMER WITH DBP.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN THE ROSTRAL
CC PORTION OF THE ANTERIOR PITUITARY DURING EMBRYOGENESIS. FOUND
CC IN SEVERAL TISSUES IN JUVENILE AND ADULT RATS.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY IN THE ANTERIOR
CC PITUITARY DURING EMBRYOGENESIS.
CC -!- SIMILARITY: BELONGS TO THE bZIP FAMILY. PAR SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; S58745; AAB20032.1; -;
DR PIR; A40579; A40579.
DR HSP; P01100; LFOS.
DR TRANSFAC; T01072;
DR InterPro; IPR001871; bZIP.
DR Pfam; PF00170; bZIP; 1.
DR SMART; SM00338; BRIZ; 1.
KW DNA-binding; Transcription regulation; Activator; Nuclear protein.
FT DOMAIN 124 175 PRO-RICH (PROLINE/ACIDIC REGION (PAR)).
FT DNA_BIND 206 215 BASIC MOTIF.
FT DOMAIN 226 240 LEUCINE-ZIPPER.
FT MUTAGEN 178 178 K->A: DECREASES AFFINITY FOR PROLACTIN
FT MUTAGEN 179 179 K->A: DECREASES AFFINITY FOR PROLACTIN
FT MUTAGEN 181 181 K->A: DECREASES AFFINITY FOR PROLACTIN
FT MUTAGEN 182 182 K->A: DECREASES AFFINITY FOR PROLACTIN
FT MUTAGEN 226 226 L->V: DIMINISHES DNA BINDING.
FT MUTAGEN 233 233 L->V: DIMINISHES DNA BINDING.
SQ SEQUENCE 261 AA; 23247 MW; 0FD3B2537AD83C4E CRC64;

Query Match 68.3%; Score 28; DB 1; Length 261;
Best Local Similarity 55.6%; Pred. No. 97;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 75 LAQNLLLPV 83

RESULT 42
YAC7_YEAST
ID YAC7_YEAST STANDARD; PRT; 261 AA.
AC P39735;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)

DE Hypothetical 29.8 kDa protein in MYO4-DRS2 intergenic region.
GN YAL027W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N.,
RA Hall J., Ouellette B.F., Keng T., Barton A.B., Su Y., Davies C.K.,
RA Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces
RT cerevisiae";
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U12980; AAC05005.1; -;
DR SGD; S0000025; YAL027W.
DR KW Hypothetical protein.
SQ SEQUENCE 261 AA; 29769 MW; F525EF2E5347E500 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 261;
Best Local Similarity 55.6%; Pred. No. 97;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 10 IARDMVLPL 18

RESULT 43
YE94_XYLFA
ID YE94_XYLFA STANDARD; PRT; 281 AA.
AC Q9PDR5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein XF1494.
GN XF1494.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Reinach F.C., Arruda P.E., Abreu F.A., Acencio M.,
RA Alvaranga R., Alves L.M.C., Araya J.E., Baia G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carrer H.,
RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Facinani L.P., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hoheisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.V., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,

RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawaaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa";
RL Nature 406:151-159(2000).
CC -!- SIMILARITY: BELONGS TO THE UPF0162 FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: AE003979; AAF84303.1; ALT_INIT.
DR InterPro; IPR001440; TPR.
DR Pfam; PF00515; TPR; 2.
DR SMART; SM00028; TPR; 2.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 281 AA; 32189 MW; 47DE879D22181684 CRC64;

Query Match 68.3%; Score 28; DB 1; Length 281;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9

DB 1 MVEQLLLPL 9

RESULT 44

CD68_MOUSE
ID CD68_MOUSE STANDARD; PRT; 326 AA.
AC P31996;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Macrosialin precursor (CD68 antigen).
GN CD68.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C; TISSUE=Macrophage;
RX MEDLINE=93252841; PubMed=8486654;
RA Holness C.L., da Silva R.P., Fawcett J., Gordon S., Simmons D.L.;
RT "Macrosialin, a mouse macrophage-restricted glycoprotein, is a member
RT of the lamp/lrp family".
RL J. Biol. Chem. 268:9661-9666(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=98148094; PubMed=9479000;
RA Li A.C., Guidez F.R.B., Collier J.G., Glass C.K.;
RT "The macrosialin promoter directs high levels of transcriptional
RT activity in macrophages dependent on combinatorial interactions
RT between PU.1 and c-Jun".
RL J. Biol. Chem. 273:5389-5399(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=98317533; PubMed=9653646;
RA Jiang Z., Shih D.M., Xia Y.R., Lusis A.J., de Beer F.C.,
RA de Villiers W.J.S., van der Westhuyzen D.R., de Beer M.C.;
RT "Structure, organization, and chromosomal mapping of the gene encoding

RT macrosialin, a macrophage-restricted protein.";
RL Genomics 50:199-205(1998).
CC -!- FUNCTION: COULD PLAY A ROLE IN PHAGOCYTIC ACTIVITIES OF TISSUE
CC MACROPHAGES, BOTH IN INTRACELLULAR LYSOSOMAL METABOLISM AND
CC EXTRACELLULAR CELL-CELL AND CELL-PATHOGEN INTERACTIONS.
CC BIND TO TISSUE- AND ORGAN-SPECIFIC LECTINS OR SELECTINS, ALLOWING
CC HOMING OF MACROPHAGE SUBSETS TO PARTICULAR SITES. RAPID
CC RECIRCULATION OF CD68 FROM ENDOSOMES, LYSOSOMES TO THE PLASMA
CC MEMBRANE MAY ALLOW MACROPHAGES TO CRAWL OVER SELECTIN BEARING
CC SUBSTRATES OR OTHER CELLS
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ENDOSOMAL OR
CC LYSOSOMAL (LONG VARIANT) AND TO A LESSER EXTENT ON THE CELL
CC SURFACE (SHORT VARIANT).
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A LONG FORM (SHOWN HERE) AND A
CC SHORT FORM; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN TISSUE MACROPHAGES AND TO A
CC LESSER EXTENT IN DENDRITIC CELLS.
CC -!- PTM: N- AND O-GLYCOSYLATED.
CC -!- SIMILARITY: BELONGS TO THE LAMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X68273; CAA48334.1; -.
DR EMBL; AF039399; AAC40056.1; -.
DR EMBL; AF022651; AAC40151.1; -.
DR PIR; S28587; S28587.
DR PIR; A46676; A46676.
DR MGD; MGI:88342; Gd68.
DR InterPro; IPR002000; Lamp.
DR Pfam; PF01299; Lamp; 1.
DR PRINTS; PR00336; LYSASOCTDMP.
DR PROSITE; PS00311; LAMP_2; 1.
KW Transmembrane; Glycoprotein; Signal; Lysosome; Repeat; Antigen;
KW Alternative splicing.
FT SIGNAL 1 20
FT CHAIN 21 326
FT DOMAIN 21 291
FT TRANSMEM 292 316
FT DOMAIN 317 326
FT DOMAIN 21 109
FT DOMAIN 110 121
FT REPEAT 44 49
FT REPEAT 50 64
FT REPEAT 65 72
FT REPEAT 73 88
FT DISULFID 139 177
FT DISULFID 249 286
FT CARBOHYD 60 60
FT CARBOHYD 84 84
FT CARBOHYD 97 97
FT CARBOHYD 129 129
FT CARBOHYD 134 134
FT CARBOHYD 169 169
FT CARBOHYD 218 218
FT CARBOHYD 233 233
FT CARBOHYD 251 251
FT VARSPPLIC 319 326
SQ SEQUENCE 326 AA; 34818 MW; AB7203A9A7E47BA CRC64;
Query Match 68.3%; Score 28; DB 1; Length 326;
Best Local Similarity 85.7%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLLPL 9

DB 290 QSLLLPL 296

```

RESULT 45
YON2 CAEEL
ID YON2 CAEEL STANDARD; PRT; 394 AA.
AC Q09523;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical 43.9 kDa GTP-binding protein E02H1.2 in chromosome II.
GN E02H1.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Smith A., Berks M.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE ERA/TRME FAMILY OF GTP-BINDING
CC PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z47075; CAA87384.1; -
DR EMBL; Z36752; CAA87384.1; JOINED.
DR EMBL; Z36752; CAA85330.1; -
DR EMBL; Z47075; CAA85330.1; JOINED.
DR WormPep: E02H1.2; CE01537.
KW Hypothetical protein; GTP-binding.
FT NP_BIND 40 47 GTP (BY SIMILARITY).
FT NP_BIND 87 91 GTP (BY SIMILARITY).
FT NP_BIND 160 163 GTP (BY SIMILARITY).
SQ SEQUENCE 394 AA; 43892 MW; 6521C28D32952BA9 CRC64;

Query Match 58.3%; Score 28; DB 1; Length 394;
Best Local Similarity 77.8%; Pred. No. 1.5e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 166 MRRSOLLPL 174

```

Search completed: November 6, 2002, 12:09:19
Job time : 8.33333 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM*protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 11 Seconds
(without alignments)
78.619 Million cell updates/sec

Title: US-09-905-083-86
Perfect score: 41
Sequence: 1 MARSLLLPL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	253	A53968	serine proteinase
2	34	82.9	461	A46394	suppressor protein
3	33	80.5	397	B87343	conserved hypothet
4	33	80.5	571	E96550	hypothetical prote
5	33	80.5	582	I48673	matrix metalloprot
6	33	80.5	582	I84471	matrix metalloprot
7	32	78.0	81	B97856	hypothetical prote
8	32	78.0	126	C82169	DNA-binding protel
9	32	78.0	432	A83060	hypothetical prote
10	32	78.0	506	B87102	conserved membrane
11	32	78.0	921	F51136	ionotropic glutama
12	32	78.0	923	F84732	probable ligand-qa
13	32	78.0	1628	T38055	hypothetical prote
14	31	75.6	138	I48107	sodium channel alp
15	31	75.6	147	S28698	hypothetical prote
16	31	75.6	169	B29016	cell division inh
17	31	75.6	169	AB0627	cell division inh
18	31	75.6	294	B86450	hypothetical prote
19	31	75.6	346	T11364	NADH dehydrogenase
20	31	75.6	375	C71917	probable transamin
21	31	75.6	375	H64597	probable transamin
22	31	75.6	452	A83734	PTS system, n-acet
23	31	75.6	491	JC6197	stromelysin 3 (EC
24	31	75.6	556	S31330	inulinase (EC 3.2.
25	31	75.6	708	T43109	cytolysin B transp
26	31	75.6	2005	B25019	sodium channel pro
27	31	75.6	5069	T17464	rifamycin polyketi
28	30	73.2	72	JC2384	corazonin precurs
29	30	73.2	74	T17834	hypothetical prote

30	73.2	170	2	S43476	histone-like DNA-b
31	30	73.2	216	JEO297	DRE/CRT-binding pr
32	30	73.2	216	T51830	transcription fact
33	30	73.2	222	D82132	hypothetical prote
34	30	73.2	222	D82132	hypothetical prote
35	30	73.2	299	T17832	transcription acti
36	30	73.2	314	D85294	transcription acti
37	30	73.2	323	A91179	probable transcript
38	30	73.2	323	S47741	probable transcript
39	30	73.2	323	B86025	probable transcript
40	30	73.2	325	A88452	protein ZC155.4 [l
41	30	73.2	325	T25122	hypothetical prote
42	30	73.2	325	A97482	hypothetical prote
43	30	73.2	325	AI2699	conserved hypothet
44	30	73.2	331	G87494	hypothetical prote
45	30	73.2	354	A48931	transmembrane glyc

ALIGNMENTS

RESULT 1

A53968
serine proteinase SCCE precursor - human
N:Alternate names: stratum corneum chymotryptic enzyme
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
C:Accession: A53968
R:Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, J. Biol. Chem. 269, 19420-19426, 1994
A:Title: Cloning, expression, and characterization of stratum corneum chymotryptic en
A:Reference number: A53968; MUID:94308225
A:Accession: A53968
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-253 <HAN>
A:Cross-references: GB:L33404; NID:G521214; NID:AAC37551.1; PID:G532504
C:Genetics:
A:Gene: GDB:PRS56; SCCE
A:Cross-references: GDB:377730
A:Map position: 7q35-7q35
C:Superfamily: trypsin; trypsin homology
F:30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 41; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.34; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0;

QY 1 MARSLLLPL 9
|||||||
Db 1 MARSLLLPL 9

RESULT 2

A46394
suppressor protein SSL1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L1531; protein YLR005W
C:Species: Saccharomyces cerevisiae
C:Date: 18-May-1994 #sequence_revision 19-Jul-1996 #text_change 20-Jun-2000
C:Accession: A46394; S64827
R:Yoon, H.; Miller, S.P.; Pabich, E.K.; Donahue, T.F.
Genes Dev. 6, 2463-2477, 1992
A:Title: SSL1, a suppressor of a HIS4 5'-UTR stem-loop mutation, is essential for tra
A:Reference number: A46394; MUID:94040711
A:Accession: A46394
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-461 <VOO>
A:Cross-references: GB:217385; NID:92695; PID:92696
R:Vandenbol, M.; Portetelle, D.; Hilger, F.
submitted to the Protein Sequence Database, May 1996
A:Reference number: S64742
A:Accession: S64827

A:Molecule type: DNA
A:Residues: 1-461 <VAN>
A:Cross-references: EMBL:Z73177; NID:gl360293; PID:gl360294; MIPS:YLR005W
A:Note: experimental_source strain S288C
C:Genetics:
A:Gene: SGB:SSL1
A:Cross-references: SGD:S0003995; MIPS:YLR005W
A:Map position: 12R
C:Keywords: transmembrane protein
F:356-372/Domain: transmembrane #status predicted <TM>

Query Match 82.9%; Score 34; DB 2; Length 461;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
||| |||||
Db 216 MARGLLLPV 224

RESULT 3
B87343
conserved hypothetical protein CC0757 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 20-Apr-2001
C:Accession: B87343
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.;
n. J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolon
n. J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: B87343
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-397 <STO>
A:Cross-references: GB:AE005673; NID:gl3421992; PIDN:AAK22742.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC0757

Query Match 80.5%; Score 33; DB 2; Length 397;
Best Local Similarity 66.7%; Pred. No. 29;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
||| |||||
Db 285 LARALLPL 293

RESULT 4
E96550
hypothetical protein Film15.13 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E96550
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719
A:Accession: E96550
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-571 <STO>
A:Cross-references: GB:AE005173; NID:g4836937; PIDN:AAD30639.1; GSPDB:GN00141
C:Genetics:
A:Gene: Film15.13

A:Map position: 1

Query Match 80.5%; Score 33; DB 2; Length 571;
Best Local Similarity 77.8%; Pred. No. 42;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
||| |||||
Db 332 MLRSLLVPL 340

RESULT 5
I48673
matrix metalloproteinase (EC 3.4.24.-) membrane type precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 22-Jun-1999
C:Accession: I48673
R:Okada, A.; Bellocq, J.P.; Rouyer, N.; Chenard, M.P.; Rio, M.C.; Chambon, P.; Basset
Proc. Natl. Acad. Sci. U.S.A. 92, 2730-2734, 1995
A:Title: Membrane-type matrix metalloproteinase (MT-MMP) gene is expressed in stromal
A:Reference number: I38046; MUID:95224014
A:Accession: I48673
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-582 <RES>
A:Cross-references: EMBL:X83536; NID:g804999; PIDN:CAA58520.1; PID:g805000
C:Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metallopro
C:Keywords: hydrolase; metalloproteinase; zinc; zymogen
F:1-23/Domain: signal sequence #status predicted <Sig>
F:24-97/Domain: activation peptide #status predicted <PRO>
F:61-284/Domain: matrix metalloproteinase homology <MMP>
F:313-508/Domain: hemopexin repeat homology <PXN>
F:93-239, 243, 249/Binding site: zinc, catalytic (His, His, His) (inhibited) #stat
F:239, 243, 249/Binding site: zinc, catalytic (His) (active) #status predicted
F:240/Active site: Glu #status predicted

Query Match 80.5%; Score 33; DB 2; Length 582;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
:|||||||
Db 8 SRSLLLPL 15

RESULT 6
I84471
matrix metalloproteinase (EC 3.4.24.-) membrane type precursor - rat
N:Alternate names: membrane-type metalloproteinase
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 22-Jun-1999
C:Accession: I84471; I61946
R:Okada, A.; Bellocq, J.P.; Rouyer, N.; Chenard, M.P.; Rio, M.C.; Chambon, P.; Basset
Proc. Natl. Acad. Sci. U.S.A. 92, 2730-2734, 1995
A:Title: Membrane-type matrix metalloproteinase (MT-MMP) gene is expressed in stromal
A:Reference number: I38046; MUID:95224014
A:Accession: I84471
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-582 <RES>
A:Cross-references: EMBL:X83537; NID:g805012; PIDN:CAA58521.1; PID:g805013
A:Accession: I61946
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-57, 'M', 69-254, 'A', 256-582 <RE2>
A:Cross-references: EMBL:X91785; NID:g1001926; PIDN:CAA62897.1; PID:g1001927
C:Genetics:
A:Gene: mt-mmp
C:Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metallopro
C:Keywords: hydrolase; metalloproteinase; zinc; zymogen
F:1-23/Domain: signal sequence #status predicted <Sig>
F:24-97/Domain: activation peptide #status predicted <PRO>
F:61-284/Domain: matrix metalloproteinase homology <MMP>

F:313-508/Domain: hemopexin repeat homology <PXN>
F:93,239,243/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #status
F:239,243,249/Binding site: zinc, catalytic (His) (active) #status predicted
F:240/Active site: Glu #status predicted

Query Match 80.5%; Score 33; DB 2; Length 582;
Best Local Similarity 87.5%; Pred. No. 43;
Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLPL 9
:|||||
Db 8 SRSLLPL 15

RESULT 7

B97856
Hypothetical protein RC1250 [imported] - Rickettsia conorii (strain Mallish 7)
C:Species: Rickettsia conorii
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 30-Sep-2001
C:Accession: B97856
R:Ogata, H.; Audic, S.; Renesto-Audiffren, P.; Fournier, P.E.; Barbe, V.; Samson, D.; R
Science 293; 2093-2098, 2001
A:Title: Mechanisms of Evolution in Rickettsia conorii and Rickettsia prowazekii.
A:Reference number: A97700; MUID:21442074; PMID:11557893
A:Accession: B97856
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-81 <KUR>
A:Cross-references: GB:AE006914; PIDN:AAL03788.1; PID:gl5620385; GSPDB:GN00173
C:Genetics:
A:Gene: RC1250

Query Match 78.0%; Score 32; DB 2; Length 81;
Best Local Similarity 75.0%; Pred. No. 9;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLPL 9
:|||||
Db 43 ARSLLPI 50

RESULT 8

C82169
DNA-binding protein inhibitor Id-2-related protein VC1696 [imported] - Vibrio cholerae
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: C82169
R:Heidelber, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, B
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: C82169
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-126 <HEI>
A:Cross-references: GB:AE004247; GB:AE003852; NID:g9656204; PIDN:AAF94846.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1696
A:Map position: 1

Query Match 78.0%; Score 32; DB 2; Length 126;
Best Local Similarity 87.5%; Pred. No. 14;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLPL 8
:|||||
Db 33 MARSLLPL 40

RESULT 9

AB3060
Hypothetical protein PA4684 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: AB3060
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: AB3060
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-432 <STO>
A:Cross-references: GB:AE004882; GB:AE004091; NID:g9950939; PIDN:AAG08071.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4684

Query Match 78.0%; Score 32; DB 2; Length 432;
Best Local Similarity 87.5%; Pred. No. 52;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLPL 9
:|||||
Db 245 ARELLPL 252

RESULT 10

B87102
conserved membrane protein ML1544 [imported] - Mycobacterium leprae
C:Species: Mycobacterium leprae
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: B87102
R:Coile, S.T.; Eiglmeier, K.; Parkhill, J.; James, K.D.; Thomson, N.R.; Wheeler, P.R.;
R.; Davies, R.M.; Devlin, K.; Duthoy, S.; Feltwell, T.; Fraser, A.; Hamlin, N.; Holro
eam, M.A.; Rutherford, K.M.
Nature 409, 1007-1011, 2001
A:Authors: Rutter, S.; Seeger, K.; Simon, S.; Simmonds, M.; Skelton, J.; Squares, R.;
A:Title: Massive gene decay in the leprosy bacillus.
A:Reference number: AB6909; MUID:21128732; PMID:11234002
A:Accession: B87102
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-506 <STO>
A:Cross-references: GB:AL450380; NID:gl30933364; PIDN:CAC30495.1; GSPDB:GN00147
C:Genetics:
A:Gene: ML1544
C:Superfamily: Mycobacterium tuberculosis hypothetical protein Rv1782

Query Match 78.0%; Score 32; DB 2; Length 506;
Best Local Similarity 77.8%; Pred. No. 62;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLPL 9
:|||||
Db 223 MARSLLPL 231

RESULT 11

T51136
ionotropic glutamate receptor glr5 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 28-Jul-2000 #sequence_revision 28-Jul-2000 #text_change 28-Jul-2000
C:Accession: T51136
R:Davenport, R.J.; Kiegle, E.A.; Tester, M.
submitted to the EMBL Data Library, December 1999
A:Description: GLR5, an ionotropic glutamate receptor ortholog from Arabidopsis.
A:Reference number: Z25309
A:Accession: T51136
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA

A:Residues: 1-921 <DAV>
A:Cross-references: EMBL:AF210701; PIDN:AAF21042.1
A:Experimental source: cultivar Columbus
C:Genetics:
A:Gene: glr5
A:Map position: 2

Query Match 78.0%; Score 32; DB 2; Length 921;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLPL 9
|||||
Db 705 MARSRLVPL 713

RESULT 12
F84732
probable ligand-gated ion channel subunit [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84732
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: F84732
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-923 <STO>
A:Cross-references: GB:AE002093; NID:g3831456; PIDN:AAC69938.1; GSPDB:GNO0139
C:Genetics:
A:Gene: At2g32400
A:Map position: 2

Query Match 78.0%; Score 32; DB 2; Length 923;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLPL 9
|||||
Db 707 MARSRLVPL 715

RESULT 13
T38055
hypothetical protein SPAC22F3.14c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T38055; T38177; S62429
R:Lye, G.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL data Library, February 1995
A:Reference number: Z21765
A:Accession: T38055
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-26 <LY>
A:Cross-references: EMBL:769239; PIDN:CAA93223.1; GSPDB:GNO00066; SPDB:SPAC1D4.14
R:Lye, G.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.
submitted to the EMBL data Library, October 1995
A:Reference number: Z21776
A:Accession: T38177
A:Molecule type: DNA
A:Residues: 8-1621 <LY>
A:Cross-references: EMBL:Z54285; NID:g1008429; PIDN:CAA91079.2; GSPDB:GNO00066; SPDB:SPAC1D4.14
C:Genetics:
A:Gene: SPAC1D4.14
A:Map position: 1L

Query Match 78.0%; Score 32; DB 2; Length 1628;
Best Local Similarity 100.0%; Pred. No. 2.1e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 RSLLLPL 9
|||||
Db 583 RSLLLPL 589

RESULT 14
T48107
sodium channel alpha subunit - long-tailed hamster (fragment)
C:Species: Cricetus longicaudatus (long-tailed hamster)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 20-Aug-1999
C:Accession: I48107
R:Lalik, P.H.; Krafe, D.S.; Ciccarelli, R.B.
Am. J. Physiol. 264, 803-809, 1993
A:Title: Characterization of endogenous Sodium channel gene expressed in chinese hams
A:Reference number: I48107
A:Accession: I48107
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-138 <RES>
A:Cross-references: GB:M87540; NID:g191067; PIDN:AAA36978.1; PID:g191068
C:Genetics:
A:Gene: chol
C:Superfamily: sodium channel protein
C:Keywords: duplication

Query Match 75.6%; Score 31; DB 2; Length 138;
Best Local Similarity 75.0%; Pred. No. 26;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLPL 8
|||||
Db 1 MARSVLVP 8

RESULT 15
S28698
hypothetical protein 16 - Agrobacterium tumefaciens plasmid pTII5955
C:Species: Agrobacterium tumefaciens
C:Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 08-Oct-1999
C:Accession: S28698
R:Barker, R.F.; Idler, K.B.; Thompson, D.V.; Kemp, J.D.
Plant Mol. Biol. 2, 335-350, 1983
A:Title: Nucleotide sequence of the T-DNA region from the Agrobacterium tumefaciens
A:Reference number: S28683
A:Accession: S28698
A:Status: translation not shown
A:Molecule type: DNA
A:Residues: 1-147 <BAR>
A:Cross-references: EMBL:X00493; NID:g39062; PIDN:CAA25178.1; PID:g39078
C:Genetics:
A:Genome: plasmid

Query Match 75.6%; Score 31; DB 2; Length 147;
Best Local Similarity 87.5%; Pred. No. 28;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLPL 8
|||||
Db 1 MARYLLPL 8

RESULT 16
B29016
cell division inhibitor sula - Salmonella typhimurium
C:Species: Salmonella typhimurium
C:Date: 05-Jun-1988 #sequence_revision 05-Jun-1988 #text_change 01-Feb-2002
C:Accession: B29016
R:Freudl, R.; Braun, G.; Honore, N.; Cole, S.T.
Gene 52, 31-40, 1987
A:Title: Evolution of the enterobacterial sula gene: a component of the SOS system en
A:Reference number: A29016; MUID:87248093


```
Db 129 AKSLMLPL 136
RESULT 21
H64597
probable transaminase (EC 2.6.1.1-) HP0624 [similarity] - Helicobacter pylori (strain 266
C:Species: Helicobacter pylori
C:Date: 30-Jan-1998 #sequence_revision 30-Jan-1998 #text_change 21-Jul-2000
C:Accession: H64597
R:Tomb, J.F.; White, O.; Kerlavage, A.R.; Clayton, R.A.; Sutton, G.G.; Fleischmann, R.D.;
Peterson, S.; Loftus, B.; Richardson, D.; Dodson, R.; Khalak, H.G.; Glodek, A.; McKenna
son, J.D.; Kelley, J.M.; Cotton, M.D.; Weidman, J.M.; Fujii, C.; Bowman, C.; Watthey, L.
Nature 388, 539-547, 1997
A:Authors: Wallin, E.; Hayes, W.S.; Borodovsky, M.; Karpk, P.D.; Smith, H.O.; Fraser, C.
A:Title: The complete genome sequence of the gastric pathogen Helicobacter pylori.
A:Reference number: A64520; MUID:97394467
A:Accession: H64597
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-375 <TOM>
A:Cross-references: GB:AE000576; GB:AE000511; NID:g2313736; PIDN:AAD07687.1; PID:g231374
C:Superfamily: aspartate transaminase
C:Keywords: aminotransferase
Query Match 75.6%; Score 31; DB 2; Length 375;
Best Local Similarity 75.0%; Pred. No. 74;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 ARSLMLPL 9
Db 129 AKSLMLPL 136
RESULT 22
A83734
PTS system, n-acetylglucosamine-specific enzyme II, ABC component (ETIABC-Nag) BH0673 [1
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
C:Accession: A83734
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20512582; PMID:11058132
A:Accession: A83734
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-452 <STO>
A:Cross-references: GB:AP001509; GB:BA000004; NID:g10173176; PIDN:BA04392.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
C:Gene: BH0673
C:Superfamily: phosphotransferase system glucose-specific enzyme II, factor II; phosphot
Query Match 75.6%; Score 31; DB 2; Length 452;
Best Local Similarity 66.7%; Pred. No. 90;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 MARSLMLPL 9
Db 8 IGRSLMLPL 16
RESULT 23
JC6197
stromelysin 3 (EC 3.4.24.-) - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 17-Mar-1999
C:Accession: JC6197
R:Okada, A.; Saez, S.; Misumi, Y.; Basset, P.
Gene 185, 187-193, 1997
A:Title: Rat stromelysin 3: cDNA cloning from healing skin wound, activation by furin an
A:Reference number: JC6197; MUID:97208872
A:Contents: Skin wounds
A:Accession: JC6197
A:Molecule type: mRNA
A:Residues: 1-491 <OKA>
A:Cross-references: GB:U46034
C:Comment: This protein is a member of the matrix metalloproteinase family.
C:Superfamily: interstitial collagenase; hemopexin repeat homology; matrix metallopro
C:Keywords: hydrolase; metalloproteinase; zinc; zymogen
F:52-261/Domain: matrix metalloproteinase homology <MMP>
F:291-483/Domain: hemopexin repeat homology <PNB>
F:84, 218, 222, 228/Binding site: zinc, catalytic (Cys, His, His) (inhibited) #stat
F:218, 222, 228/Binding site: zinc, catalytic (His) (active) #status predicted
F:219/Active site: Glu #status predicted
Query Match 75.6%; Score 31; DB 2; Length 491;
Best Local Similarity 66.7%; Pred. No. 98;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 MARSLMLPL 9
Db 11 ISRALMLPL 19
RESULT 24
S31330
inulinase (EC 3.2.1.7) - yeast (Kluyveromyces marxianus)
C:Species: Kluyveromyces marxianus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: S31330
R:Bergkamp, R.J.M.; Mooren, A.T.A.
Submitted to the EMBL Data Library, September 1992
A:Reference number: S31330
A:Accession: S31330
A:Molecule type: DNA
A:Residues: 1-556 <BER>
A:Cross-references: EMBL:X68479; NID:g2911; PIDN:CAA48500.1; PID:g2912
C:Superfamily: beta-fructofuranosidase
C:Keywords: glycosidase; hydrolase; polysaccharide degradation
Query Match 75.6%; Score 31; DB 1; Length 556;
Best Local Similarity 77.8%; Pred. No. 11e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MARSLMLPL 9
Db 3 LAYSLMLPL 11
RESULT 25
T43109
cytolysin B transport protein - Lactococcus lactis plasmid pMRC01
C:Species: Lactococcus lactis
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 02-Jun-2000
C:Accession: T43109
R:Dougherty, B.A.; Hill, C.; Weidman, J.F.; Richardson, D.R.; Venter, J.C.; Ross, R.P
Mol. Microbiol. 29, 1029-1038, 1998
A:Title: Sequence and analysis of the 60 kb conjugative, bacteriocin-producing plasm
A:Reference number: 222314
A:Accession: T43109
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-708 <DOU>
A:Cross-references: EMBL:AE001272; PIDN:AAC56012.1
A:Experimental source: strain DPC3147
C:Genetics:
A:Genome: plasmid pMRC01
A:Note: ORF00038
Query Match 75.6%; Score 31; DB 2; Length 708;
Best Local Similarity 77.8%; Pred. No. 1.4e+02;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MARSLMLPL 9
:| | | | | | |
```

Db 172 IASSLLPL 180

RESULT 26
B5019
sodium channel protein II - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Jun-1988 #sequence_revision 30-Jun-1988 #text_change 20-Aug-1999
C:Accession: B25019; S24804
R:Noda, M.; Ikeda, T.; Kayano, T.; Suzuki, H.; Takeshima, H.; Kurasaki, M.; Takahashi, H.
Nature 320, 188-192, 1986
A:Title: Existence of distinct sodium channel messenger RNAs in rat brain.
A:Reference number: A93377; MUID:86146901
A:Accession: B25019
A:Molecule type: mRNA
A:Residues: 1-2005 <NOD>
A:Experimental source: brain
R:Srao, R.; Gupta, S.K.; Auld, V.J.; Dunn, R.J.
submitted to the EMBL Data Library, August 1991
A:Description: Developmentally regulated RNA splicing of rat brain sodium channel mRNAs.
A:Reference number: S24803
A:Accession: S24804
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 183-188, 'D', 190-305 <SAR>
A:Cross-references: EMBL:X61149; NID:g57074; PIDN:CAA43458.1; PID:g57076
C:Superfamily: sodium channel protein
C:Keywords: duplication; ion transport; sodium channel; transmembrane protein; voltage-g

Query Match 75.6%; Score 31; DB 2; Length 2005;
Best Local Similarity 75.0%; Pred. No. 4.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLPL 8
||||:|
Db 1 MARSVLP 8

RESULT 27
T17464
rifamycin polyketide synthase modules 4-6 - Amycolatopsis mediterranei
C:Species: Amycolatopsis mediterranei
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 17-Nov-2000
C:Accession: T17464
R:Schupp, T.
submitted to the EMBL Data Library, December 1997
A:Reference number: Z18802
A:Accession: T17464
A>Status: Preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-5069 <SCH>
A:Cross-references: EMBL:AJ223012; NID:e1227119; PID:e1227121; PIDN:CAAL1036.1
A:Experimental source: strain LBG A3136
C:Superfamily: acyl carrier protein homology
C:Keywords: carrier protein
F:1631-1702/Domain: acyl carrier protein homology <ACP1>
F:3238-3309/Domain: acyl carrier protein homology <ACP2>
F:4939-5010/Domain: acyl carrier protein homology <ACP3>

Query Match 75.6%; Score 31; DB 2; Length 5069;
Best Local Similarity 66.7%; Pred. No. 1.1e+03;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLPL 9
|| ||||:
Db 4880 MAESLLVPI 4888

RESULT 28
JC2384
corazonin precursor - fruit fly (Drosophila melanogaster)
N:Contains: corazonin-precursor related peptide
C:Species: Drosophila melanogaster

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 29-Jan-1999
C:Accession: JC2384
R:Veestra, J.A.
Biochem. Biophys. Res. Commun. 204, 292-296, 1994
A:Title: Isolation and structure of the Drosophila corazonin gene.
A:Reference number: JC2384; MUID:95032110
A:Accession: JC2384
A:Molecule type: DNA
A:Residues: 1-72 <VEE>
A:Cross-references: GB:S74038; NID:g765231; PID:g765232
C:Genetics:
A:Gene: FlyBase:Crz
A:Cross-references: FlyBase:FBgn0013767
C:Keywords: amidated carboxyl end; pyroglutamic acid
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-30/Product: corazonin #status predicted <MAT>
F:34-67/Product: corazonin-related peptide #status predicted <CRP>
F:30/Modified site: pyroglutamate carboxylic acid (Gln) (in mature form) #status predicted
F:30/Modified site: amidated carboxyl end (Asn) (amide in mature form from following

Query Match 73.2%; Score 30; DB 2; Length 72;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLPL 9
| | | | |
Db 1 MLRLLLPL 9

RESULT 29
T17834
hypothetical protein a335L - Chlorella virus PBCV-1
C:Species: Chlorella virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17834
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17834
A>Status: Preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-74 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96703.1
A:Experimental source: specific host Chlorella strain NC64A
C:Genetics:
A:Note: a335L

Query Match 73.2%; Score 30; DB 2; Length 74;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLPL 9
: | | | | |
Db 19 LPRNLLPL 27

RESULT 30
S43476
histone-like DNA-binding protein PF 1 - oat (strain Gary)
C:Species: Avena sativa (oat)
A:Variety: Gary
C:Date: 20-Oct-1994 #sequence_revision 19-Apr-1996 #text_change 21-Jul-2000
C:Accession: S43476
R:Nieto-Sotelo, J.; Ichida, A.; Quail, P.H.
Nucleic Acids Res. 22, 1115-1116, 1994
A:Title: Positive factor 1 (PF1) from oat is an HMGY- and H1 histone-like protein tha
A:Reference number: S43476; MUID:94203798
A:Accession: S43476
A:Molecule type: mRNA
A:Residues: 1-170 <NIE>
A:Cross-references: EMBL:L24391; NID:g454278; PIDN:AAA32718.1; PID:g454279
A:Experimental source: Gary
C:Genetics:

A:Gene: PF-1
C:Superfamily: histone H1

Query Match 73.2%; Score 30; DB 2; Length 170;
Best Local Similarity 75.0%; Pred. No. 53;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 8
: : : : :
Db 6 VAKSLLLPL 13

RESULT 31

JE0297
DRE/CRT-binding protein DREB1A - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)
C:Date: 05-Dec-1998 #sequence_revision 05-Dec-1998 #text_change 02-Mar-2001
C:Accession: JE0297

R:Shinwari, Z.K.; Nakashima, K.; Miura, S.; Kasuga, M.; Seki, M.; Yamaguchi-Shinozaki, K.
Biochem. Biophys. Res. Commun. 250, 161-170, 1998

A:Title: An Arabidopsis gene family encoding DRE/CRT binding proteins involved in low-temperature stress

A:Reference number: JE0297; MUID:98407909

A:Accession: JE0297

A:Molecule type: mRNA

A:Residues: 1-216 <SHI>

A:Cross-references: DDBJ:AB013815; NID:g3660547; PIDN:BAA33434.1; PID:g3660548

C:Superfamily: Arabidopsis thaliana transcription factor DREB1B

Query Match 73.2%; Score 30; DB 2; Length 216;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
: : : : :
Db 185 MAEGMLLPL 193

RESULT 32

T51830

transcription factor DREB1A [imported] - Arabidopsis thaliana

N:Alternate names: DRE binding protein 1A

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 02-Mar-2001

C:Accession: T51830

R:Liu, Q.; Kasuga, M.; Sakuma, Y.; Abe, H.; Miura, S.; Yamaguchi-Shinozaki, K.; Shinozaki, K.
Plant Cell 10, 1391-1406, 1998

A:Title: Two transcription factors, DREB1 and DREB2, with an EREBP/AP2 DNA binding domain, respectively, in Arabidopsis.

A:Reference number: Z25477

A:Accession: T51830

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-216 <LIU>

A:Cross-references: EMBL:AB007787; PIDN:BAA33791.1

A:Experimental source: strain Colombia

C:Superfamily: Arabidopsis thaliana transcription factor DREB1B

Query Match 73.2%; Score 30; DB 2; Length 216;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
: : : : :
Db 185 MAEGMLLPL 193

RESULT 33

D82132

hypothetical protein VC1981 [imported] - Vibrio cholerae (strain N16961 serogroup O1)

C:Species: Vibrio cholerae

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001

C:Accession: D82132

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;

chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qln, H.; Dragol, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: AG2035; MUID:20406833

A:Accession: D82132

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-222 <HEI>

A:Cross-references: GB:AE004273; GB:AE003852; NID:g9656517; PIDN:AAF95129.1; GSPDB:GN

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VC1981

A:Map position: 1

Query Match 73.2%; Score 30; DB 2; Length 222;
Best Local Similarity 75.0%; Pred. No. 70;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
: : : : :
Db 111 ARQLLPL 118

RESULT 34

TI7832

hypothetical protein A333L - Chlorella virus PBCV-1

C:Species: Chlorella virus PBCV-1

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: TI7832

R:Graves, M.V.; Van Etten, J.L.

Submitted to the EMBL Data Library, May 1999

A:Reference number: Z18806

A:Accession: TI7832

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-299 <GRA>

A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96701.1

A:Experimental source: specific host Chlorella strain NC64A

C:Genetics:

A:Note: A333L

Query Match 73.2%; Score 30; DB 2; Length 299;
Best Local Similarity 66.7%; Pred. No. 96;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
: : : : :
Db 245 LPRNLLPL 253

RESULT 35

D85294

transcription activator CBFI-like protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C:Accession: D85294

R:Anonymous, The European Union Arabidopsis Genome Sequencing Consortium, The Cold Sp

Nature 402, 769-777, 1999

A:Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A:Reference number: A85001; MUID:20083488

A:Accession: D85294

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-314 <STO>

A:Cross-references: GB:NC_001268; NID:g7269398; PIDN:CAB81358.1; GSPDB:GN00140

C:Genetics:

A:Gene: AT4g25480

A:Map position: 4

Query Match 73.2%; Score 30; DB 2; Length 314;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLLPL 9
II :||||
Db 185 MAEGMLLPL 193

RESULT 36

T05799
transcription activator CBFI homolog M7J2.150 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 13-Aug-1999

R:Bevan, M.; Bargues, M.; Perez-Perez, A.; Terol, J.; Torres, A.; Perez-Alonso, M.; Hohmann
submitted to the Protein Sequence Database, April 1998
A:Reference number: T05799

A:Accession: T05799

A:Molecule type: DNA

A:Residues: 1-314 <BEV>

A:Cross-references: EMBL:AL022197

A:Experimental source: cultivar Columbia; BAC clone M7J2

C:Genetics:

A:Map position: 4

A:Introns: 215/1; 232/2

A:Note: M7J2.150

Query Match 73.2%; Score 30; DB 2; Length 314;

Best Local Similarity 66.7%; Pred. No. 1e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLLPL 9

II :||||

Db 185 MAEGMLLPL 193

RESULT 37

A91179
probable transcription regulator LYSR-type [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli

C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001

C:Accession: A91179

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hatcori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001

A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gen
A:Reference number: A99629; MUID:21156231; PMID:11258796

A:Accession: A91179

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-323 <HAY>

A:Cross-references: GB:BA000007; PIDN:BA037824.1; PID:q13363875; GSPDB:GN00154

A:Experimental source: strain O157:H7, substrain RIMD 0509952

C:Genetics:

A:Gene: ECS4401

C:Superfamily: conserved hypothetical protein H11364

Query Match 73.2%; Score 30; DB 2; Length 323;

Best Local Similarity 75.0%; Pred. No. 1e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLLPL 8

II :||||

Db 126 IAKSLLLLPL 133

RESULT 38

S47741
probable transcription regulator (tref-kdgk intergenic region) - Escherichia coli
N:Alternate names: hypothetical protein o323

C:Species: Escherichia coli

C:Date: 27-Jan-1995 #sequence_revision 27-Jan-1995 #text_change 15-Oct-1999

C:Accession: S47741; D65150

R:Plunkett, G.

submitted to the EMBL Data Library, March 1994

A:Reference number: S47666

A:Accession: S47741

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-323 <PLU>

A:Cross-references: EMBL:U00039; NID:q466582; PIDN:AAB18497.1; PID:q46658

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: D65150

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-323 <BLAT>

A:Cross-references: GB:AE000428; GB:U00096; NID:q1789931; PIDN:AAC76546.1; PID:q17899

A:Experimental source: strain K-12, substrain MGI655

C:Genetics:

A:Gene: yhjC

C:Superfamily: conserved hypothetical protein H11364

Query Match 73.2%; Score 30; DB 2; Length 323;

Best Local Similarity 75.0%; Pred. No. 1e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLLPL 8

II :||||

Db 126 IAKSLLLLPL 133

RESULT 39

B86025

probable transcription regulator LYSR-type yhjC [imported] - Escherichia coli (strain

C:Species: Escherichia coli

C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001

C:Accession: B86025

R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; May

iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apoda

Nature 409, 529-533, 2001

A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: A85480; MUID:21074935; PMID:11206551

A:Accession: B86025

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-323 <STO>

A:Cross-references: GB:AE005174; NID:q12518233; PIDN:AGS58662.1; GSPDB:GN00145; UWGP:

A:Experimental source: strain O157:H7, substrain EDU933

C:Genetics:

A:Gene: yhjC

C:Superfamily: conserved hypothetical protein H11364

Query Match 73.2%; Score 30; DB 2; Length 323;

Best Local Similarity 75.0%; Pred. No. 1e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLLPL 8

II :||||

Db 126 IAKSLLLLPL 133

RESULT 40

A88452

protein ZC155.4 [imported] - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001

C:Accession: A88452

R:anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A:Title: Genome sequence of the nematode C. elegans: a platform for investigating bio

A:Reference number: A75000; MUID:99069613; PMID:9851916

A:Note: see websites genome.wustl.edu/gsc/C_elegans/ and www.sanger.ac.uk/Projects/C-

A:Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;

A:Accession: A88452

```
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-325 <STO>
A:CROSS-references: GB:chr_III; PIDN:AA052706.1; PID:g1945571; GSPDB:GN00021; CESP:ZC155
C:Genetics:
A:Gene: ZC155.4
A:Map position: 3

Query Match      73.2%; Score 30; DB 2; Length 325;
Best Local Similarity 62.5%; Pred. No. le+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLP 8
Db 256 LARSLIMP 263
      :||||:|

RESULT 41
hypothetical protein T22G5.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25122
R:Smey, R.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19984
A:Accession: T25122
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-325 <WIL>
A:CROSS-references: EMBL:Z81127; PIDN:CAB03386.1; GSPDB:GN00023; CESP:T22G5.1
A:Experimental source: clone T22G5
C:Genetics:
A:Gene: CESP.T22G5.1
A:Map position: 5
A:Introns: 121/1; 280/2

Query Match      73.2%; Score 30; DB 2; Length 325;
Best Local Similarity 62.5%; Pred. No. le+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLP 8
Db 256 LARSLIMP 263
      :||||:|

RESULT 42
hypothetical protein AGR_C_1840 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: A97482
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quorollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughcy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97482
A:Accession: A97482
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-325 <KUR>
A:CROSS-references: GB:AE007869; PIDN:AAK86810.1; PID:g15156018; GSPDB:GN00169
C:Genetics:
A:Gene: AGR_C_1840
A:Map position: circular chromosome

Query Match      73.2%; Score 30; DB 2; Length 325;
Best Local Similarity 87.5%; Pred. No. le+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ARSLLLPL 9
Db 121 ARSLKLPL 128
      |||||

RESULT 43
conserved hypothetical protein Atul001 [imported] - Agrobacterium tumefaciens (strain
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AI2699
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kuttyavin, T.; Levy, R.; Li, M.; McCl
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB2577; PMID:11743193
A:Accession: AI2699
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-325 <KUR>
A:CROSS-references: GB:AE008688; PIDN:AAL42015.1; PID:g17739390; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atul001
A:Map position: circular chromosome

Query Match      73.2%; Score 30; DB 2; Length 325;
Best Local Similarity 87.5%; Pred. No. le+02;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ARSLLLPL 9
Db 121 ARSLKLPL 128
      |||||

RESULT 44
hypothetical protein CC1980 [imported] - Caulobacter crescentus
C:Species: Caulobacter crescentus
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: G87494
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg,
B.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Ko
n, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete Genome Sequence of Caulobacter crescentus.
A:Reference number: A87249; MUID:21173698; PMID:11259647
A:Accession: G87494
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <STO>
A:CROSS-references: GB:AE005673; NID:g13423445; PIDN:AAK23955.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC1980
C:Superfamily: conserved hypothetical protein HI0176

Query Match      73.2%; Score 30; DB 2; Length 331;
Best Local Similarity 85.7%; Pred. No. le+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 ARSLLLP 8
Db 281 ARSILLP 287
      |||||

RESULT 45
transmembrane glycoprotein CD68, 110K - human
C:Species: Homo sapiens (man)
C:Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 17-Mar-2000
C:Accession: A48931
R:Holness, C.L.; Simmons, D.L.
Blood 81, 1607-1613, 1993
```

A:Title: Molecular cloning of CD68, a human macrophage marker related to lysosomal glyco
A:Reference number: A48931; MUID:93200523
A:Accession: A48931
A>Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-354 <HOL>
A:Cross-references: GB:S57235; NID:g298664; PIDN:AAB25811.1; PID:g298665
A:Experimental source: promonocyte cell line U937
A>Note: sequence extracted from NCBI backbone (NCBIN:127492, NCBIPI:127493)
C:Superfamily: lysosome-associated membrane protein
C:Keywords: glycoprotein; transmembrane protein

Query Match 73.2%; Score 30; DB 2; Length 354;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLLPL 9
||:||||
DB 318 RSLLLPL 324

Search completed: November 6, 2002, 12:08:22
Job time : 14 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 8.11111 Seconds
(without alignments)
27.102 Million cell updates/sec

Title: US-09-905-083-86
Perfect score: 41
Sequence: 1 MARSLLLPL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/3A_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	% Match	Length	ID	Description
1	41	100.0	9	4	US-09-502-600-86
2	41	100.0	253	2	US-08-557-146-2
3	41	100.0	253	2	US-08-824-874-3
4	41	100.0	253	2	US-09-154-344-2
5	41	100.0	253	3	US-08-930-188-2
6	41	100.0	253	4	US-09-210-084-3
7	41	100.0	253	5	PCT-US96-04294-2
8	36	87.8	9	4	US-09-502-600-116
9	35	85.4	518	3	US-08-999-723-2
10	35	85.4	518	4	US-09-434-427-2
11	32	78.0	303	4	US-08-985-950-2
12	31	75.6	223	1	US-08-353-751-1
13	31	75.6	2005	3	US-08-836-325-7
14	29	70.7	23	1	US-08-353-751-2
15	29	70.7	385	4	US-09-036-987A-19
16	29	70.7	385	4	US-09-370-700-19
17	29	70.7	702	4	US-09-232-200-102
18	29	70.7	702	4	US-09-232-197-102
19	29	70.7	702	4	US-09-232-201-102
20	28	68.3	230	2	US-08-637-759B-457
21	28	68.3	230	3	US-08-871-355A-457
22	28	68.3	230	4	US-09-201-945-457
23	28	68.3	289	5	PCT-US92-00282-13
24	28	68.3	513	2	US-08-480-229C-14
25	28	68.3	513	2	US-08-659-235C-14
26	28	68.3	534	5	PCT-US92-00282-4
27	28	68.3	582	4	US-08-448-489-1

28	28	68.3	582	4	US-09-211-704A-9	Sequence 9, Appli
29	28	68.3	747	3	US-09-035-648-18	Sequence 18, Appl
30	28	68.3	747	4	US-09-001-951-18	Sequence 18, Appl
31	28	68.3	933	4	US-08-764-870-14	Sequence 14, Appl
32	28	68.3	933	4	US-08-980-115-14	Sequence 14, Appl
33	28	68.3	1620	1	US-08-542-363-2	Sequence 2, Appli
34	28	68.3	1620	4	US-09-100-089-2	Sequence 2, Appli
35	28	68.3	2152	4	US-09-036-987A-3	Sequence 3, Appli
36	28	68.3	2152	4	US-09-370-700-3	Sequence 3, Appli
37	27	65.9	9	4	US-09-502-600-36	Sequence 36, Appl
38	27	65.9	34	2	US-08-468-558-9	Sequence 9, Appli
39	27	65.9	34	4	US-08-676-444-9	Sequence 9, Appli
40	27	65.9	111	4	US-09-227-357-235	Sequence 235, App
41	27	65.9	176	1	US-08-415-751-1	Sequence 1, Appli
42	27	65.9	187	3	US-08-737-248-18	Sequence 18, Appl
43	27	65.9	307	4	US-09-177-249-13	Sequence 13, Appl
44	27	65.9	319	3	US-08-832-399-2	Sequence 2, Appli
45	27	65.9	319	4	US-09-372-498-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-09-502-600-86
; Sequence 86, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US/09/502,600A
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 86
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 1-9 of the SCCE protein
US-09-502-600-86

Query Match 100.0%; Score 41; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 1 MARSLLLPL 9

RESULT 2
US-08-557-146-2
; Sequence 2, Application US/085571146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-557-146-2

Query Match 100.0%; Score 41; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. NO. 0.18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
| | | | | | | |
Db 1 MARSLLLPL 9

RESULT 3
US-08-824-874-3
; Sequence 3, Application US/08824874
; Patent No. 5962300
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE: Filed Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:

; LIBRARY: GenBank
; CLONE: 532504
US-08-824-874-3
Query Match 100.0%; Score 41; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. NO. 0.18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MARSLLLPL 9
| | | | | | | |
Db 1 MARSLLLPL 9

RESULT 4
US-09-154-344-2
; Sequence 2, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-154-344-2

Query Match 100.0%; Score 41; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. NO. 0.18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
| | | | | | | |
Db 1 MARSLLLPL 9

RESULT 5
US-08-930-188-2
; Sequence 2, Application US/08930188
; Patent No. 6093397
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.

```

; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,188
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-930-188-2

Query Match 100.0%; Score 41; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1 MARSLLLPL 9
Db 1 MARSLLLPL 9

RESULT 6
US-09-210-084-3
; Sequence 3, Application US/09210084
; Patent No. 6197511
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,084
; FILING DATE:

```

```

; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 532504
; US-09-210-084-3

Query Match 100.0%; Score 41; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.18;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Caps 0;

Qy 1 MARSLLLPL 9
Db 1 MARSLLLPL 9

RESULT 7
PCF-US96-04294-2
; Sequence 2, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

```

; MOLECULE TYPE: protein
PCT-US96-04294-2

Query Match 100.0%; Score 41; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.18; Mismatches 0; Indels 0; Gaps 0;
Matches 9; Conservative 0;

QY 1 MARSLLLPL 9
 | | | | | | | |
Db 1 MARSLLLPL 9

RESULT 8

US-09-502-600-116
; Sequence 116, Application US/09502600A
; Patent No. 6294344

GENERAL INFORMATION:

; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; TITLE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 116
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 2-10 of the SCCE protein
US-09-502-600-116

Query Match 87.8%; Score 36; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
 | | | | | | | |
Db 1 ARSLLLPL 8

RESULT 9

US-08-999-723-2
; Sequence 2, Application US/08999723A
; Patent No. 6035180

GENERAL INFORMATION:

; APPLICANT: Powell, David J.
; APPLICANT: Southan, Christopher
; APPLICANT: Chapman, Conrad G.
; APPLICANT: Evans, Joanne R.
; TITLE OF INVENTION: ASP1
; FILE REFERENCE: GH70262
; CURRENT APPLICATION NUMBER: US/08/999,723A
; CURRENT FILING DATE: 1997-10-06
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-999-723-2

Query Match 85.4%; Score 35; DB 3; Length 518;
Best Local Similarity 77.8%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
 | | | | | | | |
Db 4 LARALLPL 12

RESULT 10

US-09-434-427-2
; Sequence 2, Application US/09434427
; Patent No. 6162630

GENERAL INFORMATION:

; APPLICANT: POWELL, DAVID J.
; APPLICANT: SOUTHAN, CHRISTOPHER
; APPLICANT: CHAPMAN, CONRAD G.
; APPLICANT: EVANS, JOANNE R.
; TITLE OF INVENTION: ASP1
; FILE REFERENCE: GH-70262-D1
; CURRENT APPLICATION NUMBER: US/09/434,427
; CURRENT FILING DATE: 1999-11-04
; EARLIER APPLICATION NUMBER: US 08/999,723
; EARLIER FILING DATE: 1997-10-06
; EARLIER APPLICATION NUMBER: UK 9626022.9
; EARLIER FILING DATE: 1996-12-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 518
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-434-427-2

Query Match 85.4%; Score 35; DB 4; Length 518;
Best Local Similarity 77.8%; Pred. No. 8;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
 | | | | | | | |
Db 4 LARALLPL 12

RESULT 11

US-08-985-950-2
; Sequence 2, Application US/08985950
; Patent No. 6140076

GENERAL INFORMATION:

; APPLICANT: Adema, Gosse Jan
; TITLE OF INVENTION: Isolated Mammalian Monocyte Cell Genes;
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,950
; FILING DATE: 05-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/041,279
; FILING DATE: 21-MARCH-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/033,181
; FILING DATE: 16-DEC-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/032,252
; FILING DATE: 06-DEC-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0670K
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650)852-9196

TELEFAX: (650)496-1204
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 303 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-985-950-2

Query Match 78.0%; Score 32; DB 4; Length 303;
Best Local Similarity 77.8%; Pred. No. 20;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLPL 9
| | | | |
DB 1 MGRPLPL 9

RESULT 12
US-08-353-751-1
; Sequence 1, Application US/08353751
; Patent No. 5712113
; GENERAL INFORMATION:
; APPLICANT: CHUNG, Bong Hyun
; APPLICANT: NAM, Soo Wan
; APPLICANT: KIM, Byung Moon
; APPLICANT: YANG, Sun Ah
; APPLICANT: PARK, Young Hoon
; TITLE OF INVENTION: SIGNAL SEQUENCES FOR SECRETION
; TITLE OF INVENTION: OF HETEROLOGOUS PROTEINS FROM
; TITLE OF INVENTION: YEAST
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CHUNG, Bong Hyun
; STREET: Hanwool Apt. 101-702, Shinsung-dong,
; CITY: Yuseong-gu
; STATE: Daejeon
; COUNTRY: Republic of Korea
; ZIP: 305-345
; ADDRESSEE: NAM, Soo Wan
; STREET: Hanbit Apt. 115-1305, 99, Oun-dong,
; CITY: Yuseong-gu
; STATE: Daejeon
; COUNTRY: Republic of Korea
; ZIP: 305-333
; ADDRESSEE: KIM, Byung Moon
; STREET: 53-2, Bokdae-dong
; CITY: Cheongju-si
; STATE: Chungcheongbuk-do
; COUNTRY: Republic of Korea
; ZIP: 360-270
; ADDRESSEE: YANG, Sun Ah
; STREET: 626-10, Shincheon 1-dong,
; CITY: Dong-gu
; STATE: Daejeon
; COUNTRY: Republic of Korea
; ZIP: 701-021
; ADDRESSEE: PARK, Young Hoon
; STREET: Town House 5-101, 391, Doryong-dong,
; CITY: Yuseong-gu
; STATE: Daejeon
; COUNTRY: Republic of Korea
; ZIP: 305-340
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 5.25inch, 1.6MB storage
; COMPUTER: IBM PC/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/353,751
FILING DATE: 12-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: KR 93-27269
FILING DATE: 10-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Shahan Islam
REGISTRATION NUMBER: 32,507
REFERENCE/DOCKET NUMBER: A-9883
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-1000
TELEFAX: (212) 953-7249
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
OTHER INFORMATION: from the 1st to the 23rd
OTHER INFORMATION: amino acids of the protein
OTHER INFORMATION: encoded in the inulinase genes
US-08-353-751-1
Query Match 75.6%; Score 31; DB 1; Length 23;
Best Local Similarity 77.8%; Pred. No. 1.8;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MARSLPL 9
| | | | |
DB 3 LAYSLPL 11
RESULT 13
US-08-836-325-7
; Sequence 7, Application US/08836325
; Patent No. 6110672
; GENERAL INFORMATION:
; APPLICANT: Mandel, Gail
; APPLICANT: Halegoua, Simon
; APPLICANT: Borden, Laurence A.
; TITLE OF INVENTION: Peripheral Nervous System Specific
; TITLE OF INVENTION: Sodium Channels, DNA Encoding Therefor, Crystallization,
; TITLE OF INVENTION: X-ray Diffraction, Computer Molecular Modeling, Rational
; TITLE OF INVENTION: Drug Design, Drug Screening, and Methods of Making and Usin
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C
; STREET: 1100 New York Ave., N. W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/836,325
; FILING DATE: 2-MAY-1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/14251
; FILING DATE: 02-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/482,401
; FILING DATE: 07-JUN-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/334,029

FILING DATE: 02-NOV-1994
ATTORNEY/AGENT INFORMATION:
NAME: Ludwig, Steven R.
REGISTRATION NUMBER: 36,203
REFERENCE/DOCKET NUMBER: 0917.024.0002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2005 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-836-325-7

Query Match 75.6%; Score 31; DB 3; Length 2005;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLP 8
| | | | : | : |
DB 1 MARSVLVP 8

RESULT 14
US-08-353-751-2
Sequence 2, Application US/08353751
Patent No. 5712113
GENERAL INFORMATION:
APPLICANT: CHUNG, Bong Hyun
APPLICANT: NAM, Soo Wan
APPLICANT: KIM, Byung Moon
APPLICANT: YANG, Sun Ah
APPLICANT: PARK, Young Hoon
TITLE OF INVENTION: SIGNAL SEQUENCES FOR SECRETION
TITLE OF INVENTION: OF HETEROLOGOUS PROTEINS FROM
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: CHUNG, Bong Hyun
STREET: Hanwool Apt. 101-702, Shinsung-dong,
CITY: Yuseong-gu
STATE: Daejeon
COUNTRY: Republic of Korea
ZIP: 305-345
ADDRESSEE: NAM, Soo Wan
STREET: Hanbit Apt. 115-1305, 99, Oun-dong,
CITY: Yuseong-gu
STATE: Daejeon
COUNTRY: Republic of Korea
ZIP: 305-333
ADDRESSEE: KIM, Byung Moon
STREET: 53-2, Bokdae-dong
CITY: Cheongju-si
STATE: Chungcheongbuk-do
COUNTRY: Republic of Korea
ZIP: 360-270
ADDRESSEE: YANG, Sun Ah
STREET: 626-10, Shincheon 1-dong,
CITY: Dong-gu
STATE: Daejeon
COUNTRY: Republic of Korea
ZIP: 701-021
ADDRESSEE: PARK, Young Hoon
STREET: Town House 5-101, 391, Doryong-dong,
CITY: Yuseong-gu
STATE: Daejeon

COUNTRY: Republic of Korea
ZIP: 305-340
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 5.25inch, 1.6MB storage
COMPUTER: IBM PC/AT
OPERATING SYSTEM: MS-DOS
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/353,751
FILING DATE: 12-DEC-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: KR 93-27269
FILING DATE: 10-DEC-1993
ATTORNEY/AGENT INFORMATION:
NAME: Shahan Islam
REGISTRATION NUMBER: 32,507
REFERENCE/DOCKET NUMBER: A-9883
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-1000
TELEFAX: (212) 953-7249
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
FEATURE:
OTHER INFORMATION: from the 1st to the 23rd
OTHER INFORMATION: amino acids of the protein encoded
OTHER INFORMATION: in the inulinase genes
US-08-353-751-2

Query Match 70.7%; Score 29; DB 1; Length 23;
Best Local Similarity 87.5%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLPL 9
| | | | |
DB 4 AYSLLPL 11

RESULT 15
US-09-036-987A-19
Sequence 19, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Patti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A

; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 385 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-036-987A-19

Query Match 70.7%; Score 29; DB 4; Length 385;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RSLLLPL 9
I:|||||
Db 347 RTLLPL 353

RESULT 16
US-09-370-700-19
; Sequence 19, Application US/09370700
; Patent No. 6274350
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H
; APPLICANT: Broughton, Mary C
; APPLICANT: Crawford, Kathryn P
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Treadway, Patti J
; APPLICANT: Turner, Jan R
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; FILE REFERENCE: 50489 DIV1
; CURRENT APPLICATION NUMBER: US/09/370,700
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: US 09/36987
; EARLIER FILING DATE: 1998-03-09
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 19
; LENGTH: 385
; TYPE: PRT
; ORGANISM: Saccharopolyspora spinosa
US-09-370-700-19

Query Match 70.7%; Score 29; DB 4; Length 385;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 RSLLLPL 9
I:|||||
Db 347 RTLLPL 353

RESULT 17
US-09-232-200-102
; Sequence 102, Application US/09232200A
; Patent No. 6288213
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MB

; CURRENT APPLICATION NUMBER: US/09/232.200A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 102
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-200-102

Query Match 70.7%; Score 29; DB 4; Length 702;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
||| |||||
Db 18 MAALLLPL 26

RESULT 18
US-09-232-197-102
; Sequence 102, Application US/09232197A
; Patent No. 6300096
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS
; FILE REFERENCE: WHI97-21p3MA
; CURRENT APPLICATION NUMBER: US/09/232.197A
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/071,374
; EARLIER FILING DATE: 1998-01-15
; EARLIER APPLICATION NUMBER: 60/093,491
; EARLIER FILING DATE: 1998-07-20
; EARLIER APPLICATION NUMBER: 60/110,941
; EARLIER FILING DATE: 1998-12-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 102
; LENGTH: 702
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-232-197-102

Query Match 70.7%; Score 29; DB 4; Length 702;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
||| |||||
Db 18 MAALLLPL 26

RESULT 19
US-09-232-201-102
; Sequence 102, Application US/09232201A
; Patent No. 6348321
; GENERAL INFORMATION:
; APPLICANT: Stahl, Andreas
; APPLICANT: Hirsch, David J.
; APPLICANT: Lodish, Harvey F.
; APPLICANT: Gimeno, Ruth E.
; APPLICANT: Tartaglia, Louis A.
; TITLE OF INVENTION: FATTY ACID TRANSPORT PROTEINS

FILE REFERENCE: WH197-21p3MC
CURRENT APPLICATION NUMBER: US/09/232,201A
CURRENT FILING DATE: 1999-01-14
EARLIER FILING DATE: 60/071,374
EARLIER FILING DATE: 1998-01-15
EARLIER FILING DATE: 60/093,491
EARLIER FILING DATE: 1998-07-20
EARLIER FILING DATE: 60/110,941
EARLIER FILING DATE: 1998-12-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: FastSeq for Windows version 3.0
SEQ ID NO 102
LENGTH: 702
TYPE: PRT
ORGANISM: Homo sapiens
US-09-232-201-102

Query Match 70.7%; Score 29; DB 4; Length 702;
Best Local Similarity 77.8%; Pred. No. 2.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
Db 18 MAALLLPL 26

RESULT 20
US-08-637-759B-457
Sequence 457, Application US/08637759B
Patent No. 5876931
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/637,759B
FILING DATE: 03-MAY-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 457:
SEQUENCE CHARACTERISTICS:
LENGTH: 230 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-637-759B-457

Query Match 68.3%; Score 28; DB 2; Length 230;

Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MARSLLLPL 9
Db 20 LSLSLLLPL 28
RESULT 21
US-08-871-355A-457
Sequence 457, Application US/08871355A
Patent No. 6015669
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/871,355A
FILING DATE: 09-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB95/02875
FILING DATE: 11-DEC-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101 CON
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 457:
SEQUENCE CHARACTERISTICS:
LENGTH: 230 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-08-871-355A-457

Query Match 68.3%; Score 28; DB 3; Length 230;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
Db 20 LSLSLLLPL 28

RESULT 22
US-09-201-945-457
Sequence 457, Application US/09201945
Patent No. 6342215
GENERAL INFORMATION:
APPLICANT: David William Holden
TITLE OF INVENTION: Identification of Genes
NUMBER OF SEQUENCES: 501
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst

STREET: 2800 One Atlantic Center
STREET: 1201 West Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/201.945
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/637,759
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: RPMS 101
TELEPHONE: (404) 873-8794
TELEFAX: (404) 873-8795
INFORMATION FOR SEQ ID NO: 457:
SEQUENCE CHARACTERISTICS:
LENGTH: 230 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
US-09-201-945-457

Query Match 68.3%; Score 28; DB 4; Length 230;
Best Local Similarity 66.7%; Pred. No. 1e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
DB 20 LSLSLLLPL 28

RESULT 23
PCT-US92-00282-13
Sequence 13, Application PC/TUS9200282
GENERAL INFORMATION:
APPLICANT: OWENS, IDA S.
APPLICANT: RITTER, JOSEPH K.
TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION THEREIN.
TITLE OF INVENTION: 40
NUMBER OF SEQUENCES: 40
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARB & CUSHMAN
STREET: 1615 L STREET, N.W.
CITY: WASHINGTON
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20036-5601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US92/00282
FILING DATE: 19920110
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: SCOTT, WATSON T.
REGISTRATION NUMBER: 26581

REFERENCE/DOCKET NUMBER: 91532-PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-861-3000
TELEFAX: 202-822-0944
TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 289 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US92-00282-13

Query Match 68.3%; Score 28; DB 5; Length 289;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
DB 1 MARGLOVPL 9

RESULT 24
US-08-480-229C-14
Sequence 14, Application US/08480229C
Patent No. 5874562
GENERAL INFORMATION:
APPLICANT: Quertermous, Thomas
APPLICANT: Hogan, Brigid
APPLICANT: Snodgrass, H. Ralph
APPLICANT: Zupancic, Thomas J.
TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
TITLE OF INVENTION: CELL LOCUS-1
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: United States
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,229C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Poissant, Brian M.
REGISTRATION NUMBER: 28,462
REFERENCE/DOCKET NUMBER: 8907-0026-999
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 Pennie
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 513 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-480-229C-14

Query Match 68.3%; Score 28; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLLP 8

Db 2 RSLLLP 7
|||||

RESULT 25

US-08-659-235C-14
; Sequence 14, Application US/08659235C
; Patent No. 5877281

GENERAL INFORMATION:

; APPLICANT: Quettermous, Thomas
; APPLICANT: Hogan, Brigid
; APPLICANT: Snodgrass, H. Ralph
; APPLICANT: Zupancic, Thomas J.
; TITLE OF INVENTION: DEVELOPMENTALLY-REGULATED ENDOTHELIAL
; TITLE OF INVENTION: CELL LOCUS-1
; NUMBER OF SEQUENCES: 29
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Pennie & Edmonds LLP
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2711

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/659,235C
; FILING DATE: 05-JUN-1996
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

; NAME: Poissant, Brian M.
; REGISTRATION NUMBER: 28,462
; REFERENCE/DOCKET NUMBER: 8907-0034-999

TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-8864/9741
; TELEX: 66141 Pennie

INFORMATION FOR SEQ ID NO: 14:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 513 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-659-235C-14

Query Match 68.3%; Score 28; DB 2; Length 513;
Best Local Similarity 100.0%; Pred. No. 2.6e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLLP 8
|||||
Db 2 RSLLLP 7

RESULT 26

PCT-US92-00282-4

; Sequence 4, Application PC/TUS9200282
; GENERAL INFORMATION:

APPLICANT: OWENS, IDA S.

; APPLICANT: RITTER, JOSEPH K.

; TITLE OF INVENTION: THE GENETIC LOCUS UGT1 AND A MUTATION

; TITLE OF INVENTION: THEREIN.

; NUMBER OF SEQUENCES: 40

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CUSHMAN DABY & CUSHMAN
; STREET: 1615 L STREET, N.W.
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: U.S.A.

; ZIP: 20036-5601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US92/00282
; FILING DATE: 19920110
; CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

; NAME: SCOTT, WATSON T.
; REGISTRATION NUMBER: 26581
; REFERENCE/DOCKET NUMBER: 91532-PCT

TELECOMMUNICATION INFORMATION:

; TELEPHONE: 202-861-3000
; TELEFAX: 202-822-0944
; TELEX: 6714627 CUSH

INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:
; LENGTH: 534 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US92-00282-4

Query Match 68.3%; Score 28; DB 5; Length 534;
Best Local Similarity 66.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
|||||
Db 1 MARGLQVPL 9

RESULT 27

US-08-448-489-1

; Sequence 1, Application US/08448489
; Patent No. 6184022

GENERAL INFORMATION:

; APPLICANT: SEIKI, Motoharu

; APPLICANT: SATO, Hiroshi

; APPLICANT: SHINAGAWA, Akira

; TITLE OF INVENTION: NOVEL METALLOPROTEINASE AND ENCODING DNA THEREFOR

; FILE REFERENCE: 55-290P

; CURRENT APPLICATION NUMBER: US/08/448,489

; CURRENT FILING DATE: 1995-06-07

; NUMBER OF SEQ ID NOS: 19

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1

; LENGTH: 582

; TYPE: PRT

; ORGANISM: Homo sapiens

US-08-448-489-1

Query Match 68.3%; Score 28; DB 4; Length 582;
Best Local Similarity 75.0%; Pred. No. 3e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
:|:|||||
Db 8 SRCLLLPL 15

RESULT 28

US-09-211-704A-9

; Sequence 9, Application US/09211704A

; Patent No. 6271014

GENERAL INFORMATION:

; APPLICANT: de Saint-Vis, Blandine Marie

; APPLICANT: Fossiez, Francois

; APPLICANT: Caux, Christophe

APPLICANT: Lebecque, Serge J.E.
TITLE OF INVENTION: Mammalian Proteinases; Related Reagents
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: DNAX Research Institute
STREET: 901 California Avenue
CITY: Palo Alto
STATE: California
COUNTRY: USA
ZIP: 94304-1104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA: US/09/211,704A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 09/005,263
FILING DATE: 09-JAN-1998
ATTORNEY/AGENT INFORMATION:
NAME: Ching, Edwin P.
REGISTRATION NUMBER: 34,090
REFERENCE/DOCKET NUMBER: SF0781K
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)852-9196
TELEFAX: (650)496-1200
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 582 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-211-704A-9

Query Match 68.3%; Score 28; DB 4; Length 582;

Best Local Similarity 75.0%; Pred. No. 3e+02; Indels 0; Gaps 0;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ARSLLLPL 9

Db 8 SRCLLPL 15

RESULT 29

US-09-035-648-18
Sequence 18, Application US/09035648

Patent No. 6100031

GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL

TITLE OF INVENTION: GROWTH AND PROLIFERATION

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/035,648

FILING DATE:

CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/818,829
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Meiklejohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 07334/003001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 747 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

US-09-035-648-18

Query Match 68.3%; Score 28; DB 3; Length 747;

Best Local Similarity 66.7%; Pred. No. 3.9e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9

Db 441 IARSLHIPL 449

RESULT 30

US-09-001-951-18

Sequence 18, Application US/09001951

Patent No. 6268470

GENERAL INFORMATION:

APPLICANT: Shyjan, Andrew W.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE

TITLE OF INVENTION: DIAGNOSIS, PREVENTION, AND TREATMENT OF NEOPLASTIC CELL

TITLE OF INVENTION: GROWTH AND PROLIFERATION

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson, P.C.

STREET: 225 Franklin Street

CITY: Boston

STATE: MA

COUNTRY: US

ZIP: 02110-2804

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/001,951

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/818,829

FILING DATE: 14-MAR-1997

APPLICATION NUMBER: 60/013,438

FILING DATE: 15-MAR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Meiklejohn, Ph.D., Anita L.

REGISTRATION NUMBER: 35,283

REFERENCE/DOCKET NUMBER: 07334/003001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-542-5070

TELEFAX: 617-542-8906

TELEX: 200154

INFORMATION FOR SEQ ID NO: 18:

SEQUENCE CHARACTERISTICS:

LENGTH: 747 amino acids

TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-001-951-18

Query Match 68.3%; Score 28; DB 4; Length 747;
Best Local Similarity 66.7%; Pred. No. 3.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 441 IARSLHPL 449

RESULT 31
US-08-764-870-14
Sequence 14, Application US/08764870
Patent No. 6236946

GENERAL INFORMATION:
APPLICANT: Scanlan, Thomas S.
APPLICANT: Baxter, John D.
APPLICANT: Fletcher, Robert J.
APPLICANT: Wagner, Richard L.
APPLICANT: Kushner, Peter J.
APPLICANT: Apriletti, James W.
APPLICANT: West, Brian
TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand Binding Domains
TITLE OF INVENTION: Binding Domains
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooley Godward
STREET: Five Palo Alto Square, 3000 El Camino Real
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/764,870
FILING DATE: 13-DEC-1996
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,540
FILING DATE: 13-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,543
FILING DATE: 13-DEC-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/008,606
FILING DATE: 14-DEC-1995

ATTORNEY/AGENT INFORMATION:
NAME: Nakamura, Jackie N.
REGISTRATION NUMBER: 35,966
REFERENCE/DOCKET NUMBER: UCAL-246/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650)843-5000
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 933 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-764-870-14

Query Match 68.3%; Score 28; DB 4; Length 933;
Best Local Similarity 85.7%; Pred. No. 5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLP 8
Db 192 ARQLLLP 198

RESULT 32
US-08-980-115-14
Sequence 14, Application US/08980115
Patent No. 6266222

GENERAL INFORMATION:
APPLICANT: Scanlan, Thomas S.
APPLICANT: Baxter, John D.
APPLICANT: Fletcher, Robert J.
APPLICANT: Wagner, Richard L.
APPLICANT: Kushner, Peter J.
APPLICANT: Apriletti, James W.
APPLICANT: West, Brian L.
APPLICANT: Shiao, Andrew K.
TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS
FILE REFERENCE: UCAL-246/02US
CURRENT APPLICATION NUMBER: US/08/980,115
CURRENT FILING DATE: 1997-11-26
EARLIER APPLICATION NUMBER: 08/764,870
EARLIER FILING DATE: 1996-12-13
EARLIER APPLICATION NUMBER: 60/008,606
EARLIER FILING DATE: 1995-12-14
EARLIER APPLICATION NUMBER: 60/008,543
EARLIER FILING DATE: 1995-12-13
EARLIER APPLICATION NUMBER: 60/008,540
EARLIER FILING DATE: 1995-12-13
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 14
LENGTH: 933
TYPE: PRT
ORGANISM: Homo sapiens
NAME/KEY: DOMAIN
FEATURE:
LOCATION: (659)..(918)
OTHER INFORMATION: minimal ligand binding domain
US-08-980-115-14

Query Match 68.3%; Score 28; DB 4; Length 933;
Best Local Similarity 85.7%; Pred. No. 5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLP 8
Db 192 ARQLLLP 198

RESULT 33
US-08-542-363-2
Sequence 2, Application US/08542363
Patent No. 5770421

GENERAL INFORMATION:
APPLICANT: Morris, Stephan W.
APPLICANT: Look, A. Thomas
TITLE OF INVENTION: ALK Protein Tyrosine Kinase/Receptor and Ligands Thereof
TITLE OF INVENTION: Ligands Thereof
NUMBER OF SEQUENCES: 43
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 New York Avenue, N.W., Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/542.363
; FILING DATE: 12-OCT-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, Samuel L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0656.0400001/SLF/GKT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1620 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-542-363-2

Query Match      68.3%; Score 28; DB 1; Length 1620;
Best Local Similarity 85.7%; Pred. No. 9.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ARSLLLP 8
   || ||||
Db 68 ARDLLLP 74

RESULT 34
US-09-100-089-2
; Sequence 2, Application US/09100089
; Patent No. 6174674
; GENERAL INFORMATION:
; APPLICANT: Morris, Stephan W.
; APPLICANT: Look, A. Thomas
; TITLE OF INVENTION: ALK Protein Tyrosine Kinase/Receptor and
; TITLE OF INVENTION: Ligands Thereof
; NUMBER OF SEQUENCES: 43
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLDSTEIN & FOX P.L.L.C.
; STREET: 1100 New York Avenue, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/100,089
; FILING DATE: Herewith
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/160,861
; FILING DATE: 03-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/542.363
; FILING DATE: 12-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Fox, Samuel L.
; REGISTRATION NUMBER: 30,353
; REFERENCE/DOCKET NUMBER: 0656.0400002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1620 amino acids
; TYPE: amino acid
; STRANDEDNESS:
```

```
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-100-089-2

Query Match      68.3%; Score 28; DB 4; Length 1620;
Best Local Similarity 85.7%; Pred. No. 9.3e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 2 ARSLLLP 8
   || ||||
Db 68 ARDLLLP 74

RESULT 35
US-09-036-987A-3
; Sequence 3, Application US/09036987A
; Patent No. 6143526
; GENERAL INFORMATION:
; APPLICANT: Baltz, Richard H.
; APPLICANT: Broughton, Mary C.
; APPLICANT: Crawford, Kathryn P.
; APPLICANT: Madduri, Krishnamurthy
; APPLICANT: Merlo, Donald J.
; APPLICANT: Treadway, Patti J.
; APPLICANT: Turner, Jan R.
; APPLICANT: Waldron, Clive
; TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
; TITLE OF INVENTION: Production
; NUMBER OF SEQUENCES: 39
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dow Agrosiences LLC Patent Department
; STREET: 9330 Zionsville Road
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: USA
; ZIP: 46268
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/036.987A
; FILING DATE: 09-MAR-1998
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stuart, Donald R
; REGISTRATION NUMBER: 28,479
; REFERENCE/DOCKET NUMBER: 50,608
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (317)337-4816
; TELEFAX: (317)337-4847
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2152 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-036-987A-3.

Query Match      68.3%; Score 28; DB 4; Length 2152;
Best Local Similarity 66.7%; Pred. No. 1.3e+03;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLPL 9
   :||| :||:
Db 1917 LARSGLLP 1925

RESULT 36
US-09-370-700-3
; Sequence 3, Application US/09370700
; Patent No. 6274350
```

GENERAL INFORMATION:
APPLICANT: Baltz, Richard H
APPLICANT: Broughton, Mary C
APPLICANT: Crawford, Kathryn P
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Treadway, Patti J
APPLICANT: Turner, Jan R
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
FILE REFERENCE: 50489 DIV1
CURRENT APPLICATION NUMBER: US/09/370,700
CURRENT FILING DATE: 1999-08-09
EARLIER APPLICATION NUMBER: US 09/36987
EARLIER FILING DATE: 1998-03-09
NUMBER OF SEQ ID NOS: 39
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3
LENGTH: 2152
TYPE: PRT
ORGANISM: Saccharopolyspora spinosa
US-09-370-700-3

Query Match 58.3%; Score 28; DB 4; Length 2152;
Best Local Similarity 56.7%; Pred. No. 1.3e+03;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
:|||||:
Db 1917 LARSGLLPI 1925

RESULT 37
US-09-502-600-36
Sequence 36, Application US/09502600A
Patent No. 6294344

GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
TITLE OF INVENTION: Ovarian Cancer
FILE REFERENCE: D6223CIP-C
CURRENT FILING DATE: 2000-02-11
CURRENT APPLICATION NUMBER: US/09/502,600A
PRIOR FILING DATE: 09/039,211
PRIOR FILING DATE: 03-14-1998
NUMBER OF SEQ ID NOS: 136
SEQ ID NO 36
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: Residues 4-12 of the SCCE protein
US-09-502-600-36

Query Match 65.9%; Score 27; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 SLLPL 9
:|||||:
Db 1 SLLPL 6

RESULT 38
US-08-468-558-9
Sequence 9, Application US/08468558
Patent No. 5877280

GENERAL INFORMATION:
APPLICANT: Wetmur, James G.
TITLE OF INVENTION: Cloning and Expression of Thermostable
TITLE OF INVENTION: Muts Genes and Proteins and Uses Therefor
NUMBER OF SEQUENCES: 38
CORRESPONDENCE ADDRESS:
ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.

STREET: Two Militia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: United States of America
ZIP: 02173
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,558
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: MSM94-05
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-861-6240
TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 34 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-468-558-9

Query Match 65.9%; Score 27; DB 2; Length 34;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
:|||||:
Db 23 AEGLLLPL 30

RESULT 39
US-08-676-444-9
Sequence 9, Application US/08676444A
Patent No. 6294325
GENERAL INFORMATION:
APPLICANT: Wetmur, James G.
TITLE OF INVENTION: CLONING AND EXPRESSION OF THERMOSTABLE
TITLE OF INVENTION: MUTL GENES AND PROTEINS AND USES THEREFOR
FILE REFERENCE: MSM95-02
CURRENT APPLICATION NUMBER: US/08/676,444A
CURRENT FILING DATE: 1996-07-05
NUMBER OF SEQ ID NOS: 48
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 9
LENGTH: 34
TYPE: PRT
ORGANISM: Thermus thermophilus
US-08-676-444-9

Query Match 65.9%; Score 27; DB 4; Length 34;
Best Local Similarity 75.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
:|||||:
Db 23 AEGLLLPL 30

RESULT 40
US-09-227-357-235
Sequence 235, Application US/09227357
Patent No. 6342581
GENERAL INFORMATION:
APPLICANT: Fischer et al.

;; TITLE OF INVENTION: 123-Human Secreted Proteins
;; FILE REFERENCE: P2010P1
;; CURRENT APPLICATION NUMBER: US/09/227,357
;; CURRENT FILING DATE: 1999-01-08
;; EARLIER APPLICATION NUMBER: PCT/US98/13684
;; EARLIER FILING DATE: 1998-07-07
;; EARLIER APPLICATION NUMBER: 60/051,926
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/052,793
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,925
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,929
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/052,803
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/052,732
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,931
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,932
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,916
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,930
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,918
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,920
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/052,733
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/052,795
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,919
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/051,928
;; EARLIER FILING DATE: 1997-07-08
;; EARLIER APPLICATION NUMBER: 60/055,722
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,723
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,948
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,949
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,953
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,950
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,947
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,964
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/056,360
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,684
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,984
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/055,954
;; EARLIER FILING DATE: 1997-08-18
;; EARLIER APPLICATION NUMBER: 60/058,785
;; EARLIER FILING DATE: 1997-09-12
;; EARLIER APPLICATION NUMBER: 60/058,664
;; EARLIER FILING DATE: 1997-09-12
;; EARLIER APPLICATION NUMBER: 60/058,660
;; EARLIER FILING DATE: 1997-09-12
;; EARLIER APPLICATION NUMBER: 60/058,661
;; EARLIER FILING DATE: 1997-09-12
;; NUMBER OF SEQ ID NOS: 672
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 235

;; LENGTH: 111
;; TYPE: PRT
;; ORGANISM: Homo sapiens
US-09-227-357-235

Query Match 65.9%; Score 27; DB 4; Length 111;
Best Local Similarity 75.0%; Pred. No. 76;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 ARSLLLPL 9
: | |||||
Db 8 SRLLLLPL 15

RESULT 41
US-08-415-751-1
; Sequence 1, Application US/08415751
; Patent No. 5643772
; GENERAL INFORMATION:
; APPLICANT: PETERSEN, CAROLYN
; APPLICANT: LEECH, JAMES
; APPLICANT: NELSON, RICHARD, C.
; APPLICANT: GUT, JIRI
; TITLE OF INVENTION: POLYPEPTIDES BINDING ANTI-
; TITLE OF INVENTION: CRYPTOSPORIDIUM ANTIBODIES, DNA
; TITLE OF INVENTION: AND RNA ENCODING THEM, HYBRID
; TITLE OF INVENTION: VECTOR AND TRANSFORMED HOST AND
; TITLE OF INVENTION: METHODS FOR IMMUNOTHERAPY AND
; NUMBER OF SEQUENCES: 50
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PHILLIPS, MOORE, LEMPJO & FINLEY
; STREET: 385 Sherman Avenue, Suite 6
; CITY: Palo Alto
; STATE: California
; COUNTRY: United States of America
; ZIP: 94306-1840
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette - 3.5 inch, 1.44 Kb storage
; COMPUTER: PC
; OPERATING SYSTEM: DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/415,751
; FILING DATE: 03-APR-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/071,880
; FILING DATE: June 1, 1993
; APPLICATION NUMBER: 07/891,301
; FILING DATE: May 29, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Hana Dolezalova
; REGISTRATION NUMBER: 30,518
; REFERENCE/DOCKET NUMBER: 480.19-2 (HHD)
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-1677
; TELEFAX: (415) 324-1678
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 176 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Cryptosporidium parvum
US-08-415-751-1

Query Match 65.9%; Score 27; DB 1; Length 176;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 2 ARSLLLPL 9
1 |||||
Db 22 ANQLLLPL 29

RESULT 42
US-08-737-248-18
; Sequence 18, Application US/08737248
; Patent No. 6114305
; GENERAL INFORMATION:
; APPLICANT: Guemene, Daniel
; APPLICANT: Zadworny, David
; APPLICANT: Karatzas, Costas
; TITLE OF INVENTION: USE OF PROLACTIN FOR PREVENTING AND/OR
; TREATING BIRD BROODINESS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WEISER & ASSOCIATES
; STREET: 230 South Fifteenth Street, Suite 500
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19102
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,248
; FILING DATE: 28-APR-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR95/00576
; FILING DATE: 03-MAY-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 94/05550
; FILING DATE: 05-MAY-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Weiser, Gerard J.
; REGISTRATION NUMBER: 19,763
; REFERENCE/DOCKET NUMBER: 989.6411P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-875-8383
; TELEFAX: 215-875-8394
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 187 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-737-248-18

Query Match 65.9%; Score 27; DB 3; Length 187;
Best Local Similarity 85.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLL 7
: |||||
Db 71 LARSLLL 77

RESULT 43
US-09-177-249-13
; Sequence 13, Application US/09177249
; Patent No. 6229064
; GENERAL INFORMATION:
; APPLICANT: Fischer, Robert L.
; APPLICANT: Ohad, Nir
; APPLICANT: Kiyosue, Tomohiro
; APPLICANT: Yadegari, Ramin
; APPLICANT: Margossian, Linda

; APPLICANT: Harada, John
; APPLICANT: Goldberg, Robert B.
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Nucleic Acids That Control Seed and Fruit
; FILE REFERENCE: 023070-08612005
; CURRENT APPLICATION NUMBER: US/09/177,249
; CURRENT FILING DATE: 1998-10-22
; EARLIER APPLICATION NUMBER: US 09/071,838
; EARLIER FILING DATE: 1998-05-01
; NUMBER OF SEQ ID NOS: 324
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 307
; TYPE: PRT
; ORGANISM: Arabidopsis sp.
US-09-177-249-13

Query Match 65.9%; Score 27; DB 4; Length 307;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 SLLLPL 9
|||||
Db 60 SLLLPL 65

RESULT 44
US-08-832-399-2
; Sequence 2, Application US/08832399
; Patent No. 6008050
; GENERAL INFORMATION:
; APPLICANT: Bergsma, Derk
; APPLICANT: Shabon, Usman
; TITLE OF INVENTION: NOVEL HUMAN NEUROTENSIN RECEPTOR TYPE 2
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY:
; ZIP: 19406
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/832,399
; FILING DATE: 02-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: King, William T.
; REGISTRATION NUMBER: 30,954
; REFERENCE/DOCKET NUMBER: GH50020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-5515
; TELEFAX: 610-270-5090
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 319 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-832-399-2

Query Match 65.9%; Score 27; DB 3; Length 319;

Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ARSLLLP 8
| | | | |
Db 143 ARSLLTP 149

RESULT 45

US-09-372-498-2
; Sequence 2, Application US/09372498
; Patent No. 6166182
; GENERAL INFORMATION:
; APPLICANT: Derek J. Bergsma
; APPLICANT: Usman Shabon
; TITLE OF INVENTION: NOVEL HUMAN NEUROTENSIN RECEPTOR TYPE 2
; FILE REFERENCE: GH-50020-1
; CURRENT APPLICATION NUMBER: US/09/372.498
; CURRENT FILING DATE: 1999-08-11
; PRIOR APPLICATION NUMBER: 08/832.399
; PRIOR FILING DATE: 1997-04-02
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 319
; TYPE: PRT
; ORGANISM: HOMO SAPIENS
US-09-372-498-2

Query Match 65.9%; Score 27; DB 4; Length 319;
Best Local Similarity 85.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 ARSLLLP 8
| | | | |
Db 143 ARSLLTP 149

Search completed: November 6, 2002, 12:06:32
Job time : 10.1111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 : Search time 23.7778 Seconds
(without alignments)
42.042 Million cell updates/sec

Title: US-09-905-083-86
Perfect score: 41
Sequence: 1 MARSLLLPL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	41	100.0	9	AAE08291	Human stratum corn
2	41	100.0	136	ABG23378	Novel human diago
3	41	100.0	253	AAR67888	Human stratum corn
4	41	100.0	253	AAW05383	Human amyloid prec
5	41	100.0	257	AAAB21326	Human HSCCE. Homo
6	36	87.8	9	AAE08320	Human stratum corn
7	35	85.4	475	AAE10657	Secreted recomblina
8	35	85.4	475	AAE02609	Human secreted asp
9	35	85.4	518	AAW61362	Aspartic proteinas
10	35	85.4	518	AAV41714	Human PRO852 prote
11	35	85.4	518	AAV22239	Human CSP56, aspar

12	35	85.4	518	20	AAV13799	Human aspartyl pro
13	35	85.4	518	21	AAB44270	Human PRO852 (UNO4
14	35	85.4	518	21	AAV88424	Human aspartyl pro
15	35	85.4	518	22	AAE10628	Human aspartyl pro
16	35	85.4	518	22	AAE10656	Human-Asp 1 protei
17	35	85.4	518	22	AAU29059	Human PRO polypept
18	35	85.4	518	22	AAE06858	Human aspartyl pro
19	35	85.4	518	22	AAU06602	Human Aspartyl pro
20	35	85.4	518	22	AAU07201	Human aspartyl pro
21	35	85.4	518	22	AAE02580	Human aspartyl pro
22	35	85.4	518	22	AAE02608	Human Aspartyl pro
23	33	80.5	272	21	AAE06855	Arabidopsis thalia
24	33	80.5	272	21	AAE06855	Arabidopsis thalia
25	33	80.5	279	21	AAE06858	Arabidopsis thalia
26	33	80.5	279	21	AAE06858	Arabidopsis thalia
27	32	78.0	43	19	AAW74766	Human secreted pro
28	32	78.0	54	21	AAE081724	Human secreted pro
29	32	78.0	105	22	AAU30194	Novel human secret
30	32	78.0	157	22	AAE064397	Amino acid sequenc
31	32	78.0	175	21	AAE07446	A human monocyte-d
32	32	78.0	226	19	AAW80407	A secreted protein
33	32	78.0	226	20	AAU08015	Human LSP-1 protei
34	32	78.0	226	21	AAE07447	A human monocyte-d
35	32	78.0	227	21	AAE07445	A human monocyte-d
36	32	78.0	227	21	AAE07230	Human signal pepti
37	32	78.0	230	21	AAE07444	A human monocyte-d
38	32	78.0	231	22	ABG17241	Novel human diago
39	32	78.0	291	19	AAW63682	Human secreted pro
40	32	78.0	303	19	AAW62772	Human immunoglobul
41	32	78.0	303	21	AAE07443	A human monocyte-d
42	32	78.0	326	22	ABE12010	Human secreted pro
43	32	78.0	654	22	AAU29036	Human PRO polypept
44	32	78.0	654	22	AAE06593	Human protein havi
45	32	78.0	654	22	AAE080243	Human PRO262 prote

ALIGNMENTS

RESULT 1
AAE08291
ID AAE08291 standard; peptide; 9 AA.
XX AC AAE08291;
XX DT 01-NOV-2001 (first entry)
XX DE Human stratum corneum chymotrypsin enzyme peptide #56 (residues 1-9).
XX DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.
XX OS Homo sapiens.
XX PN WO200159158-A1.
XX PD 16-AUG-2001.
XX PF 07-FEB-2001; 2001WO-US03977.
XX PR 11-FEB-2000; 2000US-0502600.
XX PA (UYAR-) UNIV ARKANSAS.
XX PI O'brien TJ;
XX DR WPI; 2001-514676/56.
XX PT Diagnosing cancer comprises detecting stratum corneum chymotrypsin
XX enzyme -
PS Claim 25; Page 114; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.
XX
SQ Sequence 9 AA;
Query Match 100.0%; Score 41; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
Db 1 MARSLLLPL 9
RESULT 2
ABG23378
ID ABG23378 standard; Protein: 136 AA.
XX
AC ABG23378;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #23369.
XX
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI; 2001-639362/73.
DR N-PSDB; AAS87565.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX
PS Claim 20; SEQ ID No 53737; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 136 AA;
Query Match 100.0%; Score 41; DB 22; Length 136;
Best Local Similarity 100.0%; Pred. No. 0.67;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
Db 7 MARSLLLPL 15
RESULT 3
AAR67888
ID AAR67888 standard; Protein: 253 AA.
XX
AC AAR67888;
XX
DT 09-AUG-1995 (first entry)
XX
DE Human stratum corneum chymotrophic recombinant enzyme (SCCE).
XX
KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
KW callosities; keratosis pilaris; ichthyoses; eczema.
XX
OS Homo sapiens.
XX
PN WO9500651-A.
XX
PD 05-JAN-1995.
XX
PF 20-JUN-1994; 94WO-IB00166.
XX
PR 18-JUN-1993; 93DK-0000725.
XX
PA (SYMB-) SYMBICOM AB.
XX
PI Egelrud T, Hansson L;
XX
DR WPI; 1995-052088/07.
DR N-PSDB; AAQ81203.
XX
XX Nucleotide sequences encoding stratum corneum chymotryptic enzyme
PT - and related vectors, transformed cells and polypeptides, and
PT useful for treating skin disorders, e.g. acne or psoriasis, and
PT for identification of specific inhibitors.
XX
PS Disclosure: Page 97; 137pp; English.
XX
XX The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
CC and skin care products, especially to treat and prevent acne,
CC keroderma, or other hyperkeratotic conditions (e.g. callosities or
CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
CC produced recombinantly following mammal, insect, plant, or
CC microorganism transformation with plasmid pS507.
XX
SQ Sequence 253 AA;
Query Match 100.0%; Score 41; DB 16; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
Db 1 MARSLLLPL 9

RESULT 4
AAW05383
ID AAW05383 standard; Protein; 253 AA.
AC AAW05383;
DT 31-DEC-1996 (first entry)
XX Human amyloid precursor protein protease.
DE Amyloid precursor protein protease; Alzheimer's disease; diagnosis;
KW therapy.
KW Homo sapiens.
XX OS
XX PN W09631122-A1.
XX PD 10-OCT-1996.
XX PP 02-APR-1996; 96WO-US04294.
XX PR 04-APR-1995; 95US-0416257.
XX (ELIL) LILLY & CO ELI.
PA Dixon EP, Johnstone EM, Little SP;
PI WPI; 1996-464694/46.
DR N-PSDB; AAT39783.
XX New isolated human amyloid precursor protein protease - used to
PT develop prods. for the treatment or diagnosis of associated
PT conditions, esp. Alzheimer's disease
XX Claim 1; Page 44-45; 55pp; English.
XX CC Human amyloid precursor protein protease (AAW05383) is involved in
CC the processing or clearance of amyloid precursor protein to form
CC beta-amyloid peptide. Its amino acid sequence was deduced from
CC a cDNA clone (AAT39783) obtd. from a human lung library. Recombinant
CC protease can be produced in transformed or transfected prokaryotic
CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
CC used to develop products for the design and testing of cpds. useful
CC for treating or preventing conditions associated with beta-amyloid
CC peptide, esp. Alzheimer's disease.
XX SQ Sequence 253 AA;
Query Match 100.0%; Score 41; DB 17; Length 253;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
Db 1 MARSLLLPL 9
RESULT 5
AAB21326
ID AAB21326 standard; Protein; 257 AA.
XX AAB21326;
AC AAB21326;
XX 02-FEB-2001 (first entry)
XX Human HSCEE.
XX Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCEE;
KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;
KW serine protease; cytostatic; cancer; prostrate cancer.
XX

OS Homo sapiens.
XX WO200053776-A2.
XX 14-SEP-2000.
XX 09-MAR-2000; 2000WO-CA00258.
XX 11-MAR-1999; 99US-0124260.
XX 01-APR-1999; 99US-0127386.
XX 21-JUL-1999; 99US-0144919.
XX (MOUN) MOUNT SINAI HOSPITAL.
XX Yousef GM, Diamandis EP;
XX WPI; 2000-587440/55.
XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
XX protein mediated disorders, especially cancer.
XX Example 4; Fig 17; 184pp; English.
XX The present sequence is human stratum corneum chymotryptic enzyme
XX (HSCEE), a member of the kallikrein multi-gene family. Kallikreins and
XX kallikrein-like proteins are a subgroup of the serine protease enzyme
XX family. They catalyse the selective cleavage of specific polypeptide
XX precursors to release peptides with potent biological activity. Nucleic
XX acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
XX KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
XX treatment, monitoring and diagnosis of cancers, especially prostate
XX cancer. They can also be used to identify a substance that can associate
XX with or mediate the biological activity of the proteins. Antibodies can
XX be used to treat conditions mediated by the kallikrein-like proteins.
XX SQ Sequence 257 AA;
Query Match 100.0%; Score 41; DB 21; Length 257;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
Db 1 MARSLLLPL 9
RESULT 6
AAE08320
ID AAE08320 standard; peptide; 9 AA.
XX AAE08320;
AC AAE08320;
XX 01-NOV-2001 (first entry)
XX Human stratum corneum chymotrypsin enzyme peptide #85 (residues 2-10).
XX Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisenese therapy; malignant hyperplasia.
XX Homo sapiens.
XX OS
XX WO200159158-A1.
XX 16-AUG-2001.
XX 07-FEB-2001; 2001WO-US03977.
XX 11-FEB-2000; 2000US-0502600.
XX (UYAR-) UNIV ARKANSAS.
XX O'brien TJ;

XX WPI; 2001-514676/56.
 XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
 PT enzyme -
 PT
 XX
 XX Disclosure; Page 121; 127pp; English.
 PS
 XX The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.
 XX
 XX Sequence 9 AA;
 SQ

Query Match 87.8%; Score 36; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 ARSLLLPL 9
 Db 1 ARSLLLPL 8
 :|||||
 1 ARSLLLPL 8

RESULT 7
 AAE10657
 ID AAE10657 standard; Protein; 475 AA.
 XX
 AC AAE10657;
 XX
 DT 10-DEC-2001 (first entry)
 DE
 DE Secreted recombinant hu-Asp 1 with (His)6 tag and lacking TM domain.
 KW Human; aspartyl protease 1; Aspl; amyloid precursor protein; APP;
 KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;
 KW amyloid plaque; neuronal loss; proteolytic; neuroprotective.
 XX Homo sapiens.
 OS Synthetic.
 OS
 XX GB2357767-A.
 PN
 XX
 XX
 PD 04-JUL-2001.
 XX
 XX
 PF 22-SEP-2000; 2000GB-0023315.
 XX
 XX 23-SEP-1999; 99US-0155493.
 PR 23-SEP-1999; 99US-0404133.
 PR 23-SEP-1999; 99WO-US20881.
 PR 13-OCT-1999; 99US-0416901.
 PR 06-DEC-1999; 99US-0169232.
 XX
 XX (PHAA) PHARMACIA & UPJOHN CO.
 PA
 XX
 XX Blenkowski MJ, Gurney M;
 PI
 XX
 XX WPI; 2001-444208/48.
 DR
 XX

Polypeptide comprising fragments of human aspartyl protease with
 PT amyloid precursor protein processing activity and alpha-secretase
 PT activity, for identifying modulators useful in treating Alzheimer's
 PT disease -
 XX
 XX Example 14; Page 156-158; 187pp; English.
 PS The patent discloses human aspartyl protease 1 (hu-Aspl) or modified
 XX Aspl proteins which lack transmembrane domain or amino terminal
 CC

CC domain or cytoplasmic domain and retains alpha-secretase activity
 CC and amyloid protein precursor (APP) processing activity. The proteins
 CC of the invention are useful for assaying hu-Aspl alpha-secretase
 CC activity, which in turn is useful for identifying modulators of
 CC hu-Aspl alpha-secretase activity, where modulators that increase
 CC hu-Aspl alpha-secretase activity, where modulators that increase
 CC disease (AD) which causes progressive dementia with consequent
 CC formation of amyloid plaques, neurofibrillary tangles, gliosis and
 CC neuronal loss. Hu-Aspl protease substrate is useful for assaying
 CC hu-Aspl proteolytic activity, by contacting hu-Aspl protein with
 CC the substrate under acidic conditions and determining the level
 CC of hu-Aspl proteolytic activity. The present sequence is secreted
 CC recombinant human Asp 1 protein lacking a transmembrane (TM) domain
 CC and containing a (His)6 tag. This sequence is generated from human
 CC Asp 1 protein by the deletion of its C-terminal TM domain and addition
 CC of hexa-histidine tag at its C-terminus.
 XX
 XX Sequence 475 AA;
 SQ

Query Match 85.4%; Score 35; DB 22; Length 475;
 Best Local Similarity 77.8%; Pred. No. 46;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MARSLLLPL 9
 Db 4 LARALLPL 12
 :||:||||
 1 MARSLLLPL 9
 4 LARALLPL 12

RESULT 8
 AAE02609
 ID AAE02609 standard; Protein; 475 AA.
 XX
 AC AAE02609;
 XX
 DT 10-AUG-2001 (first entry)
 DE
 DE Human secreted aspartyl protease-1 (Asp-1) deltaTM (His)6 protein.
 KW Human; alpha-secretase; amyloid precursor protein; APP; therapy;
 KW Alzheimer's disease; antialzheimer's; aspartyl protease 1; Aspl;
 KW beta-secretase; secreted Asp-1 deltaTM (His)6 protein.
 XX Homo sapiens.
 OS Synthetic.
 OS
 XX WO200123533-A2.
 PN
 XX
 PD 05-APR-2001.
 XX
 XX 22-SEP-2000; 2000WO-US26080.
 PF 23-SEP-1999; 99US-0155493.
 PR 23-SEP-1999; 99WO-US20881.
 PR 13-OCT-1999; 99US-0416901.
 PR 06-DEC-1999; 99US-0169232.
 XX
 XX (PHAA) PHARMACIA & UPJOHN CO.
 PA
 XX
 XX Gurney M, Bienkowski MJ;
 PI
 XX
 XX WPI; 2001-290516/30.
 DR
 XX

Enzymes that cleave the alpha-secretase site of the amyloid precursor
 PT protein, useful for the treatment of Alzheimer's disease -
 XX
 XX Example 14; Page 184-186; 189pp; English.
 PS The present invention relates to enzymes for cleaving the alpha-
 XX secretase site of the amyloid precursor protein (APP) and methods of
 CC identifying those enzymes. The methods may be used to identify enzymes
 CC that may be used to cleave the alpha-secretase cleavage site of the APP
 CC protein. The enzymes may be used to treat or modulate the progress of
 CC Alzheimer's disease. The present sequence is human secreted recombinant

CC Aspartyl protease-1 (Asp-1) deltaTM (His)6 protein which is used for the
CC expression of pre-pro-human-Aspartyl protease 1 (Asp1). The protein is
CC obtained by replacing C-terminal transmembrane and cytoplasmic domains
CC with a hexahistidine purification tag in the human Aspartyl protease 1.
XX
SQ Sequence 475 AA;

Query Match 85.4%; Score 35; DB 22; Length 475;
Best Local Similarity 77.8%; Pred. No. 46;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
:|:|:|:|:|
Db 4 LARALLLPL 12

RESULT 9

AAW61362
ID AAW61362 standard; Protein; 518 AA.

XX
AC AAW61362;

DT 25-SEP-1998 (first entry)

XX
XX Aspartic proteinase Asp1.

XX
XX Asp1; Aspartic proteinase; Alzheimer's disease; cancer; melanoma.

XX
XX Homo sapiens.

XX
XX EP848062-A2.

XX
XX 17-JUN-1998.

XX
PF 01-DEC-1997; 97EP-0309648.

XX
PR 14-DEC-1996; 96GB-0026022.

XX
PA (SMIK) SMITHKLINE BEECHAM CORP.

XX
PA (SMIK) SMITHKLINE BEECHAM PLC.

XX
PI Chapman CG, Evans JR, Powell DJ, Southan C;

XX
XX WPI; 1998-314477/28.

DR
N-PSDB; AAV27962.

XX
PT New isolated polynucleotide encodes Aspartic protease polypeptide -
PT used to diagnosis, treat and vaccinate against Alzheimer's disease,
PT cancer and melanoma

XX
PS Claim 11; Page 7; 19pp; English.

XX
CC The human ASP1 protein is structurally related to other proteins of the
CC Aspartic proteinase family. ASP 1 polypeptides and polynucleotides can
CC be used to diagnosis, treat and vaccinate against Alzheimer's disease,
CC cancer and melanoma.

XX
SQ Sequence 518 AA;

Query Match 85.4%; Score 35; DB 19; Length 518;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
:|:|:|:|:|
Db 4 LARALLLPL 12

RESULT 10

AAAY41714
ID AAY41714 standard; Protein; 518 AA.

XX
AC AAY41714;

XX
DT 07-DEC-1999 (first entry)
XX
DE Human PRO852 protein sequence.

XX
KW Human; PRO; EST; expressed sequence tag; PCR primer; hybridisation;
KW probe; blood coagulation disorder; cancer; cellular adhesion disorder;
KW secreted protein; transmembrane protein.

OS
XX Homo sapiens.

XX
PN WO9946281-A2.

XX
PD 16-SEP-1999.

XX
XX 08-MAR-1999; 99WO-US05028.

XX
PR 10-MAR-1998; 98US-0077450.

PR
PR 11-MAR-1998; 98US-0077632.

PR
PR 11-MAR-1998; 98US-0077641.

PR
PR 11-MAR-1998; 98US-0077649.

PR
PR 12-MAR-1998; 98US-0077791.

PR
PR 13-MAR-1998; 98US-0078004.

PR
PR 17-MAR-1998; 98US-0040220.

PR
PR 20-MAR-1998; 98US-0078886.

PR
PR 20-MAR-1998; 98US-0078910.

PR
PR 20-MAR-1998; 98US-0078936.

PR
PR 20-MAR-1998; 98US-0078939.

PR
PR 25-MAR-1998; 98US-0079294.

PR
PR 26-MAR-1998; 98US-0079656.

PR
PR 27-MAR-1998; 98US-0079663.

PR
PR 27-MAR-1998; 98US-0079689.

PR
PR 27-MAR-1998; 98US-0079728.

PR
PR 27-MAR-1998; 98US-0079786.

PR
PR 30-MAR-1998; 98US-0079920.

PR
PR 31-MAR-1998; 98US-0080105.

PR
PR 31-MAR-1998; 98US-0080107.

PR
PR 31-MAR-1998; 98US-0080165.

PR
PR 31-MAR-1998; 98US-0080194.

PR
PR 01-APR-1998; 98US-0080327.

PR
PR 01-APR-1998; 98US-0080328.

PR
PR 01-APR-1998; 98US-0080333.

PR
PR 01-APR-1998; 98US-0080334.

PR
PR 08-APR-1998; 98US-0081049.

PR
PR 08-APR-1998; 98US-0081070.

PR
PR 08-APR-1998; 98US-0081071.

PR
PR 09-APR-1998; 98US-0081195.

PR
PR 09-APR-1998; 98US-0081203.

PR
PR 09-APR-1998; 98US-0081229.

PR
PR 15-APR-1998; 98US-0081817.

PR
PR 15-APR-1998; 98US-0081836.

PR
PR 15-APR-1998; 98US-0081952.

PR
PR 15-APR-1998; 98US-0081955.

PR
PR 21-APR-1998; 98US-0082568.

PR
PR 21-APR-1998; 98US-0082569.

PR
PR 22-APR-1998; 98US-0082700.

PR
PR 22-APR-1998; 98US-0082704.

PR
PR 22-APR-1998; 98US-0082804.

PR
PR 23-APR-1998; 98US-0082767.

PR
PR 23-APR-1998; 98US-0082796.

PR
PR 27-APR-1998; 98US-0083336.

PR
PR 28-APR-1998; 98US-0083322.

PR
PR 29-APR-1998; 98US-0083392.

PR
PR 29-APR-1998; 98US-0083495.

PR
PR 29-APR-1998; 98US-0083496.

PR
PR 29-APR-1998; 98US-0083499.

PR
PR 29-APR-1998; 98US-0083500.

PR
PR 29-APR-1998; 98US-0083545.

PR
PR 29-APR-1998; 98US-0083554.

PR
PR 29-APR-1998; 98US-0083558.

PR
PR 29-APR-1998; 98US-0083559.

PR 30-APR-1998; 98US-0083742.
PR 05-MAY-1998; 98US-0084366.
PR 06-MAY-1998; 98US-0084414.
PR 06-MAY-1998; 98US-0084441.
PR 07-MAY-1998; 98US-0084458.
PR 07-MAY-1998; 98US-0084598.
PR 07-MAY-1998; 98US-0084600.
PR 07-MAY-1998; 98US-0084627.
PR 07-MAY-1998; 98US-0084637.
PR 07-MAY-1998; 98US-0084639.
PR 07-MAY-1998; 98US-0084640.
PR 07-MAY-1998; 98US-0084643.
PR 13-MAY-1998; 98US-0085323.
PR 13-MAY-1998; 98US-0085328.
PR 13-MAY-1998; 98US-0085339.
PR 15-MAY-1998; 98US-0085573.
PR 15-MAY-1998; 98US-0085579.
PR 15-MAY-1998; 98US-0085580.
PR 15-MAY-1998; 98US-0085582.
PR 15-MAY-1998; 98US-0085689.
PR 15-MAY-1998; 98US-0085697.
PR 15-MAY-1998; 98US-0085700.
PR 15-MAY-1998; 98US-0085704.
PR 22-MAY-1998; 98US-0086023.
PR 22-MAY-1998; 98US-0086392.
PR 22-MAY-1998; 98US-0086414.
PR 22-MAY-1998; 98US-0086430.
PR 28-MAY-1998; 98US-0086486.
PR 28-MAY-1998; 98US-0087098.
PR 28-MAY-1998; 98US-0087106.
PR 30-JUL-1998; 98US-0087208.
PR 11-SEP-1998; 98US-0094651.
PR 11-SEP-1998; 98US-0100038.
PA (GETH) GENENTECH INC.
XX
PI Wood WI, Goddard A, Gurney A, Yuan J, Baker KP, Chen J;
XX WPI: 1999-551358/46.
XX N-PSDB; AAZ34056.
XX
PT New secreted and transmembrane polypeptides and their polynucleotides,
PT useful for treating blood coagulation disorders, cancers and cellular
PT adhesion disorders -
XX
XX
PS Claim 12; Fig 73; 530pp; English.
XX
CC The present invention describes secreted and transmembrane polypeptides
CC and their polynucleotides. The nucleotide sequences are useful as
CC sources of probes, primers, for chromosome mapping, and for generation
CC of antisense sequences. They can also be used to create transgenic
CC animals. The proteins can be used to treat a variety of diseases and
CC disorders, depending on their function. Diseases that may be treated
CC include blood coagulation disorders, cancers and cellular adhesion
CC disorders. They may also be used to raise antibodies. AAZ33891 to
CC AAZ34338, and AAY41685 to AAY41774 represent polynucleotide and
CC polypeptide sequence given in the exemplification of the present
XX invention.
XX
SQ Sequence 518 AA;
Query Match 85.4%; Score 35; DB 20; Length 518;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MARSLLLPL 9
Db :||:|||||
4 LARALLLPL 12
RESULT 11
AAY22239
ID AAY22239 standard; Protein; 518 AA.
XX

AC AAY22239;
XX
DT 20-SEP-1999 (first entry)
XX
DE Human CSP56, aspartyl-type protease, protein sequence.
XX
KW Metastatic marker protein; human; cancer metastasis; breast cancer;
KW colon cancer; diagnosis; therapy; tumour; metastatic potential;
KW CSP56; aspartyl-type protease.
XX
OS Homo sapiens.
XX
PN WO9934004-A2.
XX
PD 08-JUL-1999.
XX
PF 24-DEC-1998; 98WO-US27608.
XX
PR 31-DEC-1997; 97US-0070112.
XX
PA (CHIR) CHIRON CORP.
XX
PI Giese K, Xin H;
XX
DR WPI: 1999-430248/36.
DR N-PSDB; AAX84708.
XX
PT New polynucleotides associated with cancer metastasis
XX
PS Claim 4; Page 78-80; 80pp; English.
XX
CC This sequence represents a polypeptide of the invention, and is
CC an aspartyl-type protease, designated CSP56. The polynucleotides (PNS) of
CC the invention encode metastatic marker protein variants. The PNS and
CC polypeptides can be used as markers for cancer metastasis. The products
CC can be used for identifying metastatic tissue or metastatic potential of
CC a tissue, e.g. breast or colon tissue. They can also be used for
CC screening test compounds for the ability to suppress the metastatic
CC potential of a tumour. The products can be used for developing products
CC for the therapy of cancers, particularly breast or colon cancer.
XX
SQ Sequence 518 AA;
Query Match 85.4%; Score 35; DB 20; Length 518;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MARSLLLPL 9
Db :||:|||||
4 LARALLLPL 12
RESULT 12
AAY13799
ID AAY13799 standard; Protein; 518 AA.
XX
AC AAY13799;
XX
DT 21-SEP-1999 (first entry)
XX
DE Human aspartyl protease, CSP56.
XX
KW CSP56; human; aspartyl protease; diagnosis; neoplasia; tumour;
KW breast tumour; colon tumour.
XX
OS Homo sapiens.
XX
PN WO9933963-A1.
XX
PD 08-JUL-1999.
XX
PF 14-DEC-1998; 98WO-US26547.
XX

CC This sequence represents the human aspartyl protease amino acid sequence.
CC The invention relates to a protease capable of cleaving the beta
CC secretase site of amyloid precursor protein (APP). The protease contains
CC a sequence encoding the amino acid sequence DTG and a sequence encoding
CC DSG or DTG separated by 100-300 amino acids. When mutated the APP gene
CC causes an autosomal dominant form of Alzheimer's disease. APP localises
CC to the cell surface membrane and have a single C-terminal transmembrane
CC domain. Proteolytic processing of APP produces the amyloid beta protein,
CC which is possibly very important in Alzheimer's disease. The invention
CC includes a nucleotide sequence encoding the protease, a vector containing
CC the nucleotide sequence, and a cell line comprising the vector. Methods
CC for screening for inhibitors of beta secretase activity are also given in
CC the invention. The human aspartase protein and nucleotide sequences and
CC the methods for identifying inhibitors of the protease, are useful in the
XX treatment of and research in to Alzheimer's disease.

Query Match 85.4%; Score 35; DB 21; Length 518;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MARSLLLPL 9
Db 4 LARALLLPL 12
:||:|||||

RESULT 15
AAE10628
ID AAE10628 standard; Protein; 518 AA.
AC AAE10628;
XX
XX
DT 10-DEC-2001 (first entry)
DE Human aspartyl protease 1 (hu-Asp1) protein.
DE Human; aspartyl protease 1; Asp1; amyloid precursor protein; APP;
KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;
KW amyloid plaque; neuronal loss; proteolytic; neurotropic; neuroprotective;
XX chromosome 21.
OS Homo sapiens.
FH
FT Key Location/Qualifiers
FT Peptide 1..20
FT Protein /label= Signal_peptide
FT 21..518
FT Domain /note= "Mature human aspartyl protease 1"
FT 469..492
XX /label= Transmembrane_domain
PN GB2357767-A.
XX
XX 04-JUL-2001.
XX
XX 22-SEP-2000; 2000GB-0023315.
XX
XX 23-SEP-1999; 99US-0155493.
XX 23-SEP-1999; 99US-0404133.
PR 23-SEP-1999; 99WO-US20881.
PR 13-OCT-1999; 99US-0416901.
PR 06-DEC-1999; 99US-0169232.
XX (PHAA) PHARMACIA & UPJOHN CO.
XX
XX Bienkowski MJ, Gurney M;
XX WPI; 2001-444208/48.
XX N-PSD; AADI7864.
XX

Polypeptide comprising fragments of human aspartyl protease with
PT amyloid precursor protein processing activity and alpha-secretase
PT disease .
XX
XX Example 14; Page 155-156; 187pp; English.

PT activity, for identifying modulators useful in treating Alzheimer's
PT disease .
XX
XX Claim 36; Fig 1; 187pp; English.
XX
CC The patent discloses human aspartyl protease 1 (hu-Asp1) or modified
CC Asp1 proteins which lack transmembrane domain or amino terminal
CC domain or cytoplasmic domain and retains alpha-secretase activity
CC and amyloid protein precursor (APP) processing activity. The proteins
CC of the invention are useful for assaying hu-Asp1 alpha-secretase
CC activity, which in turn is useful for identifying modulators of
CC hu-Asp1 alpha-secretase activity, where modulators that increase
CC disease (AD) which causes progressive dementia with consequent
CC formation of amyloid plaques, neurofibrillary tangles, gliosis and
CC neuronal loss. Hu-Asp1 protease substrate is useful for assaying
CC hu-Asp1 proteolytic activity, by contacting hu-Asp1 protein with
CC the substrate under acidic conditions and determining the level of
CC hu-Asp1 proteolytic activity. The present sequence is Asp1 protein
CC from human. Asp1 gene is localised on chromosome 21.
XX
SQ Sequence 518 AA;

Query Match 85.4%; Score 35; DB 22; Length 518;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MARSLLLPL 9
Db 4 LARALLLPL 12
:||:|||||

RESULT 16
AAE10656
ID AAE10656 standard; Protein; 518 AA.
AC AAE10656;
XX
XX 10-DEC-2001 (first entry)
DE Human-Asp 1 protein lacking TM domain and containing (His)6 tag.
DE Human; aspartyl protease 1; Asp1; amyloid precursor protein; APP;
KW Alzheimer's disease; AD; dementia; neurofibrillary tangle; gliosis;
KW amyloid plaque; neuronal loss; proteolytic; neurotropic; neuroprotective.
XX Homo sapiens.
XX Synthetic.
XX GB2357767-A.
XX
XX 04-JUL-2001.
XX
XX 22-SEP-2000; 2000GB-0023315.
XX
XX 23-SEP-1999; 99US-0155493.
XX 23-SEP-1999; 99US-0404133.
PR 23-SEP-1999; 99WO-US20881.
PR 13-OCT-1999; 99US-0416901.
PR 06-DEC-1999; 99US-0169232.
XX (PHAA) PHARMACIA & UPJOHN CO.
XX
XX Bienkowski MJ, Gurney M;
XX WPI; 2001-444208/48.
XX

Polypeptide comprising fragments of human aspartyl protease with
PT amyloid precursor protein processing activity and alpha-secretase
PT activity, for identifying modulators useful in treating Alzheimer's
PT disease .
XX
XX Example 14; Page 155-156; 187pp; English.

XX The patent discloses human aspartyl protease 1 (hu-Aspl) or modified
 CC Aspl proteins which lack transmembrane domain or amino terminal
 CC domain or cytoplasmic domain and retains alpha-secretase activity
 CC and amyloid protein precursor (APP) processing activity. The proteins
 CC of the invention are useful for assaying hu-Aspl alpha-secretase
 CC activity, which in turn is useful for identifying modulators of
 CC hu-Aspl alpha-secretase activity, where modulators that increase
 CC hu-Aspl alpha-secretase activity are useful for treating Alzheimer's
 CC disease (AD) which causes progressive dementia with consequent
 CC formation of amyloid plaques, neurofibrillary tangles, gliosis and
 CC neuronal loss. Hu-Aspl protease substrate is useful for assaying
 CC hu-Aspl proteolytic activity, by contacting hu-Aspl protein with
 CC the substrate under acidic conditions and determining the level of
 CC hu-Aspl proteolytic activity. The present sequence is human Asp 1
 CC protein lacking a transmembrane (TM) domain and containing (His)6
 CC tag. This sequence is generated from human Asp 1 protein by the
 CC deletion of its C-terminal TM domain and addition of hexa-histidine
 CC tag at its C-terminus.
 XX
 SQ Sequence 518 AA;

Query Match 85.4%; Score 35; DB 22; Length 518;
 Best Local Similarity 77.8%; Pred. No. 50;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
 Db :||:|||||
 4 LARALLLPL 12

RESULT 17
 AAU29059
 ID AAU29059 standard; Protein: 518 AA.
 XX
 AC AAU29059;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Human PRO polypeptide sequence #36.
 XX
 KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep;
 KW dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha;
 KW blood; chondrocyte cell; cell proliferation; cell differentiation; colon;
 KW adrenal; lung; breast; prostate; rectum; cervix; liver; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 XX WQ200168848-A2.
 XX
 XX 20-SEP-2001.
 XX
 XX 28-FEB-2001; 2001WO-US06520.
 XX
 XX 01-MAR-2000; 2000WO-US05601.
 XX 02-MAR-2000; 2000WO-US05841.
 XX 03-MAR-2000; 2000US-187202P.
 XX 06-MAR-2000; 2000US-186988P.
 XX 14-MAR-2000; 2000US-189320P.
 XX 15-MAR-2000; 2000US-189328P.
 XX 21-MAR-2000; 2000WO-US06884.
 XX 21-MAR-2000; 2000US-190828P.
 XX 21-MAR-2000; 2000US-191007P.
 XX 21-MAR-2000; 2000US-191048P.
 XX 21-MAR-2000; 2000US-191314P.
 XX 28-MAR-2000; 2000US-192655P.
 XX 29-MAR-2000; 2000US-193032P.
 XX 29-MAR-2000; 2000US-193053P.
 XX 30-MAR-2000; 2000WO-US08439.
 XX 04-APR-2000; 2000US-194449P.
 XX 11-APR-2000; 2000US-194647P.
 XX 11-APR-2000; 2000US-195975P.
 XX 11-APR-2000; 2000US-196000P.

PR 11-APR-2000; 2000US-196187P.
 PR 11-APR-2000; 2000US-196690P.
 PR 11-APR-2000; 2000US-196820P.
 PR 18-APR-2000; 2000US-198121P.
 PR 18-APR-2000; 2000US-198585P.
 PR 25-APR-2000; 2000US-199397P.
 PR 25-APR-2000; 2000US-199550P.
 PR 25-APR-2000; 2000US-199654P.
 PR 03-MAY-2000; 2000US-201516P.
 PR 17-MAY-2000; 2000WO-US13705.
 PR 22-MAY-2000; 2000WO-US14042.
 PR 30-MAY-2000; 2000WO-US14941.
 PR 02-JUN-2000; 2000WO-US15264.
 PR 05-JUN-2000; 2000US-209832P.
 PR 28-JUL-2000; 2000WO-US20710.
 PR 22-AUG-2000; 2000US-0644848.
 PR 24-AUG-2000; 2000WO-US23328.
 PR 08-NOV-2000; 2000WO-US30952.
 PR 01-DEC-2000; 2000WO-US32678.
 PR 20-DEC-2000; 2000WO-US34956.
 XX
 XX (GETH) GENENTECH INC.
 XX
 XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;
 PI Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;
 XX
 XX WPI: 2001-602746/68.
 DR N-PSDB; AAS45960.
 XX
 PT Novel nucleic acids encoding PRO polypeptides, used to diagnose the
 PT presence of tumours, such as prostate and breast tumours, in mammals and
 PT to screen for modulators of the compounds -
 XX
 XX Claim 11: Fig 72; 774pp; English.
 XX
 CC Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention.
 CC The PRO polypeptides and their associated nucleic acids can be used to
 CC detect the presence of a tumour in a mammal by comparing the level of
 CC expression of a PRO polypeptide in a test sample of cells from the animal
 CC and a control sample of normal cells, whereby a higher level of
 CC expression in the test sample indicates the presence of a tumour in the
 CC mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats
 CC and rabbits but are preferably human. The polypeptides can be used to
 CC stimulate tumour necrosis factor (TNF) alpha release from human blood,
 CC when contacted with it. A specific polypeptide can be used to stimulate
 CC the proliferation or differentiation of chondrocyte cells. The PRO
 CC proteins can be used to determine the presence of tumours and also
 CC susceptibility to tumour development, particularly adrenal, lung, colon,
 CC breast, prostate, rectal, cervical, or liver tumours, in mammalian
 CC subjects. The oligonucleotide probes specific for the PRO nucleic acids
 CC can be used for genetic analysis of individuals with genetic disorders.
 XX
 SQ Sequence 518 AA;

Query Match 85.4%; Score 35; DB 22; Length 518;
 Best Local Similarity 77.8%; Pred. No. 50;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
 Db :||:|||||
 4 LARALLLPL 12

RESULT 18
 AAU29059
 ID AAU29059 standard; Protein: 518 AA.
 XX
 XX AAU29059;
 XX
 XX 23-OCT-2001 (first entry)
 XX
 XX Human aspartyl protease 1 (Hu-Aspl) protein.
 XX

KW Human; aspartyl protease 1; Asp 1; beta-amyloid precursor protein; APP;
 KW beta-secretase; Alzheimer's disease; dementia; amyloid plaque; gliosis;
 KW neurofibrillary tangle; neuronal loss; amyloid-beta peptide; nontropic;
 KW neuroprotective; antisense therapy; gene therapy; chromosome 21.
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH Peptide 1..20
 FT Protein /label= Signal_peptide
 FT 21..518
 FT Domain /note= "Mature human aspartyl protease 1 (Hu-Asp1)"
 FT 469..492
 FT /label= Transmembrane_domain

XX WO200150829-A2.
 PN 19-JUL-2001.
 PD 09-MAY-2001; 2001WO-IB00799.
 XX 09-MAY-2001; 2001WO-IB00799.
 PD 09-MAY-2001; 2001WO-IB00799.

XX (BIEN/) BIENKOWSKI M J.
 PA (GURN/) GURNEY M E.
 PA (HEIN/) HEINRIKSON R L.
 PA (PARO/) PARODI L A.
 PA (YANR/) YAN R.

XX BIENKOWSKI MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;
 XX WPI: 2001-483072/52.
 DR N-PSDB; AAD13020.

XX Novel purified polypeptide comprising fragment of mammalian aspartyl
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta
 PT secretase activity of Asp2 useful for identifying inhibitors of Asp2
 PT activity

XX Example 2; Fig 1; 185pp; English.

CC The invention relates to human aspartyl proteases (Hu-Asp), beta-amyloid
 CC precursor protein (APP) isoforms and their corresponding DNA molecules.
 CC Human aspartyl proteases can act as beta-secretase proteases useful for
 CC treating Alzheimer's disease. APP isoforms are useful for identifying
 CC modulators of amyloid-beta peptide production, for use in designing
 CC therapeutics for the treatment and prevention of Alzheimer's disease,
 CC dementia, formation of amyloid plaques, neurofibrillary tangles, gliosis
 CC and neuronal loss. APP isoforms are also used in methods for identifying
 CC inhibitors and modulators of human Asp2 activity. The invention relates
 CC to a method for identifying agents that modulate the activity of human
 CC aspartyl protease Asp2. Amyloid-beta peptides obtained from APP are used
 CC as a means to screen in cellular assays for the inhibitors of beta- and
 CC gamma- secretase. Hu-Asp DNA fragments are useful as probes or primers in
 CC polymerase chain reactions (PCR). The probes are useful for detecting
 CC Hu-Asp nucleic acids in in vitro assays and in Northern and Southern
 CC blots. The present sequence is human aspartyl protease 1 (Hu-Asp1).
 CC Hu-Asp 1 gene is localised on chromosome 21.

XX Sequence 518 AA;
 Query Match 85.4%; Score 35; DB 22; Length 518;
 Best Local Similarity 77.8%; Pred. No. 50;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
 Db 4 LARALLPL 12

RESULT 19
 AAU06602
 ID AAU06602 standard; Protein: 518 AA.

XX AAU06602;
 AC 24-OCT-2001 (first entry)
 XX Human Aspartyl protease 1 (Asp1).
 DE Human; Aspartyl protease; Asp1; Asp2; beta-secretase; nontropic;
 KW neuroprotective; amyloid protein precursor; APP; Alzheimer's disease;
 KW amyloid-beta; Abeta.
 XX Homo sapiens.
 OS WO200149098-A2.
 PN 12-JUL-2001.
 PD 09-MAY-2001; 2001WO-IB00799.
 XX 09-MAY-2001; 2001WO-IB00799.
 PD 09-MAY-2001; 2001WO-IB00799.
 XX (BIEN/) BIENKOWSKI M J.
 PA (GURN/) GURNEY M E.
 PA (HEIN/) HEINRIKSON R L.
 PA (PARO/) PARODI L A.
 PA (YANR/) YAN R.

XX BIENKOWSKI MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;
 XX WPI: 2001-502549/55.
 DR N-PSDB; AAS11516.

XX Novel purified polypeptide comprising fragment of mammalian aspartyl
 PT protease 2, lacking Asp2 transmembrane domain and retaining beta
 PT secretase activity of Asp2 useful for identifying inhibitors of Asp2
 PT activity

XX Example 2; Fig 1; 185pp; English.

CC The invention relates to a purified polypeptide comprising a fragment of
 CC mammalian aspartyl protease (Asp2) protein which lacks the Asp2
 CC transmembrane domain and the Asp2 protein, and where the polypeptide and
 CC the fragment retain the beta-secretase activity of the mammalian Asp2
 CC protein. The invention also details polynucleotides for the Asp
 CC proteins and vectors expressing them, and a polypeptide (isoform of
 CC amyloid protein precursor (APP)) comprising the amino acid sequence of an
 CC APP or its fragment containing an APP cleavage site recognizable by a
 CC mammalian beta-secretase, and further comprising two lysine residues at
 CC the carboxyl terminus of the amino acid sequence of the mammalian APP or
 CC APP fragment. Also included in the invention are methods of identifying
 CC modulators or inhibitors of Asp2. Modulators and inhibitors of identifying
 CC useful for treating Alzheimer's disease. APP is useful in methods for
 CC identifying inhibitors or modulators of human Asp2 activity and
 CC amyloid-beta (Abeta) peptide production. APP is also useful in designing
 CC therapeutics for the treatment or prevention of Alzheimer's disease.
 CC APP comprising the APP-Sw-beta-secretase peptide sequence (NLDA), which
 CC is associated with increased levels of Abeta processing is useful in
 CC assays relating the Alzheimer's research. The expression vector is useful
 CC for recombinantly expressing APP. Nucleic acids that hybridise to
 CC Asp oligonucleotides are useful as probes or primers. The probes are
 CC useful for detecting Hu-Asp nucleic acids in in vitro assays and in
 CC Northern and Southern blots. The present sequence is human Asp1.

XX Sequence 518 AA;
 Query Match 85.4%; Score 35; DB 22; Length 518;
 Best Local Similarity 77.8%; Pred. No. 50;
 Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
 Db 4 LARALLPL 12

RESULT 20
AAU07201
ID AAU07201 standard; Protein; 518 AA.
XX AC AAU07201;
XX DT 24-OCT-2001 (first entry)
XX XX Human aspartyl protease 1 (Asp-1).
XX DE
XX DE Human; aspartyl protease 1; Asp-1; nootropic; neuroprotective;
KW aspartyl protease 2; Asp2; amyloid protein precursor; APP;
KW beta-secretase; Alzheimer's disease.
XX OS
XX OS Homo sapiens.
XX PN WC200149097-A2.
XX PD 12-JUL-2001.
XX PF 09-MAY-2001; 2001WO-IB00797.
XX PR 09-MAY-2001; 2001WO-IB00797.
XX XX (BIEN/) BIENKOWSKI M J.
PA (GURNEY/) GURNEY M E.
PA (HEIN/) HEINRIKSON R L.
PA (PARODI/) PARODI L A.
PA (YANR/) YAN R.

XX BIENKOWSKI MJ, Gurney ME, Heinrikson RL, Parodi LA, Yan R;
XX WPI: 2001-502548/55.
XX N-PSDB; AAS11701.
XX Novel purified polypeptide comprising fragment of mammalian aspartyl
PT protease 2, lacking Asp2 transmembrane domain and retaining beta
PT secretase activity of Asp2 useful for identifying inhibitors of Asp2
PT activity
XX Example 2; Fig 1; 185pp; English.

XX The invention relates to a novel purified polypeptide comprising a
CC fragment of mammalian aspartyl protease 2 (Asp2) protein which lacks the
CC Asp2 transmembrane domain and the Asp2 protein, and where the polypeptide
CC and the fragment retain the beta-secretase activity of the mammalian Asp2
CC protein. Also included is an isoform of amyloid protein precursor (APP)
CC comprising the amino acid sequence of a APP or its fragment containing
CC an APP cleavage site recognizable by a mammalian beta-secretase, and
CC further comprising two lysine residues at the carboxyl terminus of the
CC amino acid sequence of the mammalian APP or APP fragment. The
CC polypeptides are used for assaying for modulators of beta-secretase
CC activity; identifying agents that inhibit the APP processing activity
CC of human Asp2 aspartyl protease (Hu-Asp2); identifying agents that
CC modulate the activity of Asp2; and for reducing cellular production of
CC amyloid beta (Abeta) from APP. Agents identified by the above methods
CC are useful for treating Alzheimer's disease; and for identifying
CC modulators of amyloid-beta (Abeta) peptide production, for use in
CC designing therapeutics for the treatment or prevention of Alzheimer's
CC disease. Probes and primers derived from Asp nucleic acid sequences
CC are useful for detecting Hu-Asp nucleic acids in in vitro assays and in
CC Northern and Southern blots. The present sequence represents the
CC amino acid sequence of human Asp-1.

XX Sequence 518 AA;
SQ Query Match 85.4%; Score 35; DB 22; Length 518;
Best Local Similarity 77.8%; Pred. No. 50;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
:|||||

Db 4 LARALLPL 12

RESULT 21
AAE02580
ID AAE02580 standard; Protein; 518 AA.
XX AC AAE02580;
XX DT 10-AUG-2001 (first entry)
XX DE Human aspartyl protease 1 (Asp 1).
XX DE Human; alpha-secretase; amyloid precursor protein; APP; therapy;
KW Alzheimer's disease; antialzheimer's; aspartyl protease 1; Asp 1;
KW beta-secretase; chromosome 21.
XX OS
XX OS Homo sapiens.
XX XX
XX FH Key Location/Qualifiers
FT Peptide 1..20
FT Peptide /label= Signal_peptide
FT Peptide 22..62
FT Peptide /label= Asp_1_prepropeptide
FT Peptide 23..62
FT Peptide /label= Asp_1_propeptide
FT Protein 63..518
FT /label= Mature_human_Asp_1_protein
FT /note= "Specifically claimed"
FT Active-site 87..89
FT /label= Active_site_1
FT Active-site 110..113
FT /label= Active_site_2
FT Active-site 303..305
FT /label= Active_site_3
FT Domain 469..492
FT /label= Transmembrane_domain
FT Domain 493..518
FT /label= Cytoplasmic_domain
FT Region 497..518
FT /note= "Peptide #1"
XX XX
XX PN WO2001235333-A2.
XX PD 05-APR-2001.
XX PF 22-SEP-2000; 2000WO-US26080.
XX PR 23-SEP-1999; 99US-0155493.
XX PR 23-SEP-1999; 99WO-US20881.
XX PR 13-OCT-1999; 99US-0416901.
XX PR 06-DEC-1999; 99US-0169232.
XX (PHAA) PHARMACIA & UPJOHN CO.
XX Gurney M, Bienkowski MJ;
XX WPI; 2001-290516/30.
XX N-PSDB; AAD06738.
XX Enzymes that cleave the alpha-secretase site of the amyloid precursor
PT protein, useful for the treatment of Alzheimer's disease -
XX Claim 29; Fig 1; 189pp; English.
XX The present invention relates to enzymes for cleaving the alpha-
CC secretase site of the amyloid precursor protein (APP) and methods of
CC identifying those enzymes. The methods may be used to identify enzymes
CC that may be used to cleave the alpha-secretase cleavage site of the APP
CC protein. The enzymes may be used to treat or modulate the progress of
CC Alzheimer's disease. The present sequence is human aspartyl protease 1
CC (Asp 1). Asp 1 has alpha-secretase protease and beta-secretase
CC protease activities. Asp 1 gene is located on chromosome 21.

XX SQ Sequence 518 AA; Query Match 85.4%; Score 35; DB 22; Length 518; Best Local Similarity 77.8%; Pred. No. 50; Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9 :||:||||| Db 4 LARALLPL 12

RESULT 22 AAE02608 ID AAE02608 standard; Protein: 518 AA. AC AAE02608; DT 10-AUG-2001 (first entry) DE Human Aspartyl protease-1 (Asp-1) deltaTM (His)6 protein. KW Human; alpha-secretase; amyloid precursor protein; App; therapy; Alzheimer's disease; antialzheimer's; aspartyl protease 1; Aspi; beta-secretase; Asp-1 deltaTM (His)6 protein. OS Homo sapiens. OS Synthetic. PN WO200123533-A2. PD 05-APR-2001. PF 22-SEP-2000; 200WO-US26080. PR 23-SEP-1999; 99US-0155493. PR 23-SEP-1999; 99WO-US20881. PR 13-OCT-1999; 99US-0416901. PR 06-DEC-1999; 99US-0169232. XX PA (PHAA) PHARMACIA & UPJOHN CO. XX Gurney M, Bienkowski MJ; WPI: 2001-290516/30. PT Enzymes that cleave the alpha-secretase site of the amyloid precursor protein, useful for the treatment of Alzheimer's disease - Example 14; Page 183-184; 189pp; English. CC The present invention relates to enzymes for cleaving the alpha-secretase site of the amyloid precursor protein (APP) and methods of identifying those enzymes. The methods may be used to identify enzymes that may be used to cleave the alpha-secretase cleavage site of the APP protein. The enzymes may be used to treat or modulate the progress of Alzheimer's disease. The present sequence is human Aspartyl protease-1 (Asp-1) deltaTM (His)6 protein which is used for the expression of pre-pro-human-Aspartyl protease 1 (Aspi). This protein is obtained by replacing C-terminal transmembrane and cytoplasmic domains with a hexahistidine purification tag in the human Aspartyl protease 1.

XX SQ Sequence 518 AA; Query Match 85.4%; Score 35; DB 22; Length 518; Best Local Similarity 77.8%; Pred. No. 50; Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9 :||:||||| Db 4 LARALLPL 12

RESULT 23 AAG08685 ID AAG08685 standard; Protein; 272 AA. XX AAG08685; AC AAG08685; DT 17-OCT-2000 (first entry) DE Arabidopsis thaliana protein fragment SEQ ID NO: 6321. XX Arabidopsis thaliana. KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence. OS Arabidopsis thaliana. PN EP1033405-A2. PD 06-SEP-2000. PF 25-FEB-2000; 2000EP-0301439. PR 25-FEB-1999; 99US-0121825. PR 05-MAR-1999; 99US-0123180. PR 09-MAR-1999; 99US-0123548. PR 23-MAR-1999; 99US-0125788. PR 25-MAR-1999; 99US-0126264. PR 29-MAR-1999; 99US-0126785. PR 01-APR-1999; 99US-0127462. PR 06-APR-1999; 99US-0128234. PR 08-APR-1999; 99US-0128714. PR 16-APR-1999; 99US-0129845. PR 19-APR-1999; 99US-0130077. PR 21-APR-1999; 99US-0130449. PR 23-APR-1999; 99US-0130510. PR 23-APR-1999; 99US-0130891. PR 28-APR-1999; 99US-0131449. PR 30-APR-1999; 99US-0132048. PR 30-APR-1999; 99US-0132407. PR 04-MAY-1999; 99US-0132484. PR 05-MAY-1999; 99US-0132485. PR 06-MAY-1999; 99US-0132486. PR 06-MAY-1999; 99US-0132487. PR 07-MAY-1999; 99US-0132863. PR 11-MAY-1999; 99US-0134256. PR 14-MAY-1999; 99US-0134218. PR 14-MAY-1999; 99US-0134219. PR 14-MAY-1999; 99US-0134221. PR 14-MAY-1999; 99US-0134370. PR 18-MAY-1999; 99US-0134768. PR 19-MAY-1999; 99US-0134941. PR 20-MAY-1999; 99US-0135124. PR 21-MAY-1999; 99US-0135353. PR 24-MAY-1999; 99US-0135629. PR 25-MAY-1999; 99US-0136021. PR 27-MAY-1999; 99US-0136392. PR 28-MAY-1999; 99US-0136782. PR 01-JUN-1999; 99US-0137222. PR 03-JUN-1999; 99US-0137528. PR 04-JUN-1999; 99US-0137502. PR 07-JUN-1999; 99US-0137724. PR 08-JUN-1999; 99US-0138094. PR 10-JUN-1999; 99US-0138540. PR 10-JUN-1999; 99US-0138847. PR 14-JUN-1999; 99US-0139119. PR 16-JUN-1999; 99US-0139452. PR 16-JUN-1999; 99US-0139453. PR 17-JUN-1999; 99US-0139492. PR 18-JUN-1999; 99US-0139454. PR 18-JUN-1999; 99US-0139455. PR 18-JUN-1999; 99US-0139456. PR 18-JUN-1999; 99US-0139457. PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142290.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142920.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.
PR 14-JUL-1999; 9905-0143624.
PR 15-JUL-1999; 9905-0144005.
PR 16-JUL-1999; 9905-0144085.
PR 16-JUL-1999; 9905-0144086.
PR 19-JUL-1999; 9905-0144325.
PR 19-JUL-1999; 9905-0144331.
PR 19-JUL-1999; 9905-0144332.
PR 19-JUL-1999; 9905-0144333.
PR 19-JUL-1999; 9905-0144334.
PR 19-JUL-1999; 9905-0144335.
PR 20-JUL-1999; 9905-0144352.
PR 20-JUL-1999; 9905-0144632.
PR 20-JUL-1999; 9905-0144634.
PR 21-JUL-1999; 9905-0144814.
PR 21-JUL-1999; 9905-0145086.
PR 21-JUL-1999; 9905-0145088.
PR 22-JUL-1999; 9905-0145085.
PR 22-JUL-1999; 9905-0145087.
PR 22-JUL-1999; 9905-0145089.
PR 22-JUL-1999; 9905-0145192.
PR 23-JUL-1999; 9905-0145145.
PR 23-JUL-1999; 9905-0145218.
PR 23-JUL-1999; 9905-0145224.
PR 26-JUL-1999; 9905-0145276.
PR 27-JUL-1999; 9905-0145913.
PR 27-JUL-1999; 9905-0145918.
PR 27-JUL-1999; 9905-0145919.
PR 28-JUL-1999; 9905-0145951.
PR 02-AUG-1999; 9905-0146386.
PR 02-AUG-1999; 9905-0146388.
PR 02-AUG-1999; 9905-0146389.
PR 03-AUG-1999; 9905-0147038.
PR 04-AUG-1999; 9905-0147204.
PR 04-AUG-1999; 9905-0147302.
PR 05-AUG-1999; 9905-0147192.
PR 05-AUG-1999; 9905-0147260.
PR 06-AUG-1999; 9905-0147303.
PR 06-AUG-1999; 9905-0147416.
PR 09-AUG-1999; 9905-0147493.
PR 09-AUG-1999; 9905-0147935.
PR 10-AUG-1999; 9905-0148171.
PR 11-AUG-1999; 9905-0148319.
PR 12-AUG-1999; 9905-0148341.
PR 13-AUG-1999; 9905-0148565.
PR 13-AUG-1999; 9905-0148684.
PR 16-AUG-1999; 9905-0149368.
PR 17-AUG-1999; 9905-0149175.
PR 18-AUG-1999; 9905-0149426.
PR 20-AUG-1999; 9905-0149722.
PR 20-AUG-1999; 9905-0149723.

PR 20-AUG-1999; 9905-0149929.
PR 23-AUG-1999; 9905-0149902.
PR 23-AUG-1999; 9905-0149930.
PR 25-AUG-1999; 9905-0150566.
PR 26-AUG-1999; 9905-0150884.
PR 27-AUG-1999; 9905-0151065.
PR 27-AUG-1999; 9905-0151066.
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 80.5%; Score 33; DB 21; Length 272;
Best Local Similarity 88.9%; Pred. No. 66;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9

Db 1 MARSLLLSL 9

RESULT 24

AAG45920

ID AAG45920 standard; Protein; 272 AA.

XX AAG45920;

XX AC

XX DT 18-OCT-2000 (first entry)

XX	Arabidopsis thaliana protein fragment SEQ ID NO: 57711.	PR	21-JUN-1999;	99US-0139817.
DE		PR	22-JUN-1999;	99US-0139899.
XX	Protein identification; signal transduction pathway; metabolic pathway;	PR	23-JUN-1999;	99US-0140353.
KW	hybridisation assay; genetic mapping; gene expression control; promoter;	PR	24-JUN-1999;	99US-0140354.
KW	termination sequence.	PR	25-JUN-1999;	99US-0140695.
XX		PR	28-JUN-1999;	99US-0140823.
OS	Arabidopsis thaliana.	PR	29-JUN-1999;	99US-0140991.
XX		PR	30-JUN-1999;	99US-0141287.
PN	EP1033405-A2.	PR	01-JUL-1999;	99US-0141842.
XX		PR	01-JUL-1999;	99US-0142154.
PD	06-SEP-2000.	PR	02-JUL-1999;	99US-0142055.
XX		PR	06-JUL-1999;	99US-0142390.
PF		PR	08-JUL-1999;	99US-0142803.
XX		PR	09-JUL-1999;	99US-0142920.
PR	25-FEB-2000; 2000EP-0301439.	PR	12-JUL-1999;	99US-0142977.
PR	25-FEB-1999;	PR	13-JUL-1999;	99US-0143542.
PR	05-MAR-1999;	PR	14-JUL-1999;	99US-0143624.
PR	09-MAR-1999;	PR	15-JUL-1999;	99US-0144005.
PR	23-MAR-1999;	PR	16-JUL-1999;	99US-0144085.
PR	25-MAR-1999;	PR	16-JUL-1999;	99US-0144086.
PR	29-MAR-1999;	PR	19-JUL-1999;	99US-0144325.
PR	01-APR-1999;	PR	19-JUL-1999;	99US-0144331.
PR	06-APR-1999;	PR	19-JUL-1999;	99US-0144332.
PR	08-APR-1999;	PR	19-JUL-1999;	99US-0144333.
PR	16-APR-1999;	PR	19-JUL-1999;	99US-0144334.
PR	19-APR-1999;	PR	19-JUL-1999;	99US-0144335.
PR	21-APR-1999;	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	PR	20-JUL-1999;	99US-0144632.
PR	28-APR-1999;	PR	20-JUL-1999;	99US-0144884.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	PR	21-JUL-1999;	99US-0145086.
PR	04-MAY-1999;	PR	21-JUL-1999;	99US-0145088.
PR	05-MAY-1999;	PR	22-JUL-1999;	99US-0145085.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	PR	22-JUL-1999;	99US-0145192.
PR	11-MAY-1999;	PR	23-JUL-1999;	99US-0145145.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	PR	27-JUL-1999;	99US-0145913.
PR	18-MAY-1999;	PR	27-JUL-1999;	99US-0145918.
PR	19-MAY-1999;	PR	27-JUL-1999;	99US-0145919.
PR	20-MAY-1999;	PR	28-JUL-1999;	99US-0145951.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	PR	06-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	PR	09-AUG-1999;	99US-0147935.
PR	10-JUN-1999;	PR	10-AUG-1999;	99US-0148171.
PR	14-JUN-1999;	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	PR	12-AUG-1999;	99US-0148341.
PR	16-JUN-1999;	PR	13-AUG-1999;	99US-0148565.
PR	17-JUN-1999;	PR	13-AUG-1999;	99US-0148684.
PR	18-JUN-1999;	PR	16-AUG-1999;	99US-0149368.
PR	18-JUN-1999;	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	PR	18-AUG-1999;	99US-0149426.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	PR	27-AUG-1999;	99US-0151066.


```
PR 27-AUG-1999; 9905-0151080.
PR 30-AUG-1999; 9905-0151303.
PR 31-AUG-1999; 9905-0151438.
PR 01-SEP-1999; 9905-0151930.
PR 07-SEP-1999; 9905-0152363.
PR 10-SEP-1999; 9905-0153070.
PR 13-SEP-1999; 9905-0153758.
PR 15-SEP-1999; 9905-0154018.
PR 16-SEP-1999; 9905-0154039.
PR 20-SEP-1999; 9905-0154779.
PR 22-SEP-1999; 9905-0155139.
PR 23-SEP-1999; 9905-0155486.
PR 24-SEP-1999; 9905-0155659.
PR 28-SEP-1999; 9905-0156458.
PR 29-SEP-1999; 9905-0156596.
PR 04-OCT-1999; 9905-0157117.
PR 05-OCT-1999; 9905-0157753.
PR 06-OCT-1999; 9905-0157865.
PR 07-OCT-1999; 9905-0158029.
PR 08-OCT-1999; 9905-0158232.
PR 12-OCT-1999; 9905-0158369.
PR 13-OCT-1999; 9905-0159293.
PR 13-OCT-1999; 9905-0159294.
PR 13-OCT-1999; 9905-0159295.
PR 14-OCT-1999; 9905-0159329.
PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160768.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match: 80.5%; Score 33; DB 21; Length 272;
Best Local Similarity 88.9%; Pred. No. 66;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 1 MARSLLLSL 9

RESULT 25
AAG08684
ID AAG08684 standard; Protein; 279 AA.
XX AC AAG08684;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 6320.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
```

```
OS Arabidopsis thaliana.
XX EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 9905-0121825.
PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137502.
PR 07-JUN-1999; 9905-0137724.
PR 08-JUN-1999; 9905-0138094.
PR 10-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 14-JUN-1999; 9905-0139119.
PR 16-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139763.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
```

```
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145226.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 08-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147310.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148568.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149829.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 80.5%; Score 33; DB 21; Length 279;
Best Local Similarity 88.9%; Pred. No. 67;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9
Db 8 MARSLLLSL 16

RESULT 26
AAG45919
ID AAG45919 standard; Protein; 279 AA.
XX
AC AAG45919;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 57710.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
```

XX	25-FEB-1999;	99US-0121825.
PR	05-MAR-1999;	99US-0123180.
PR	09-MAR-1999;	99US-0123548.
PR	23-MAR-1999;	99US-0125788.
PR	25-MAR-1999;	99US-0126264.
PR	29-MAR-1999;	99US-0126785.
PR	01-APR-1999;	99US-0127462.
PR	06-APR-1999;	99US-0128234.
PR	08-APR-1999;	99US-0128714.
PR	16-APR-1999;	99US-0129845.
PR	19-APR-1999;	99US-0130077.
PR	21-APR-1999;	99US-0130449.
PR	23-APR-1999;	99US-0130510.
PR	28-APR-1999;	99US-0130891.
PR	30-APR-1999;	99US-0131449.
PR	30-APR-1999;	99US-0132048.
PR	04-MAY-1999;	99US-0132407.
PR	05-MAY-1999;	99US-0132484.
PR	06-MAY-1999;	99US-0132485.
PR	06-MAY-1999;	99US-0132486.
PR	07-MAY-1999;	99US-0132487.
PR	11-MAY-1999;	99US-0132863.
PR	14-MAY-1999;	99US-0134256.
PR	14-MAY-1999;	99US-0134218.
PR	14-MAY-1999;	99US-0134219.
PR	14-MAY-1999;	99US-0134221.
PR	18-MAY-1999;	99US-0134370.
PR	18-MAY-1999;	99US-0134768.
PR	19-MAY-1999;	99US-0134941.
PR	20-MAY-1999;	99US-0135124.
PR	21-MAY-1999;	99US-0135353.
PR	24-MAY-1999;	99US-0135629.
PR	25-MAY-1999;	99US-0136021.
PR	27-MAY-1999;	99US-0136392.
PR	28-MAY-1999;	99US-0136782.
PR	01-JUN-1999;	99US-0137222.
PR	03-JUN-1999;	99US-0137528.
PR	04-JUN-1999;	99US-0137502.
PR	07-JUN-1999;	99US-0137724.
PR	08-JUN-1999;	99US-0138094.
PR	10-JUN-1999;	99US-0138540.
PR	10-JUN-1999;	99US-0138847.
PR	14-JUN-1999;	99US-0139119.
PR	16-JUN-1999;	99US-0139452.
PR	16-JUN-1999;	99US-0139453.
PR	17-JUN-1999;	99US-0139492.
PR	18-JUN-1999;	99US-0139454.
PR	18-JUN-1999;	99US-0139455.
PR	18-JUN-1999;	99US-0139456.
PR	18-JUN-1999;	99US-0139457.
PR	18-JUN-1999;	99US-0139458.
PR	18-JUN-1999;	99US-0139459.
PR	18-JUN-1999;	99US-0139460.
PR	18-JUN-1999;	99US-0139461.
PR	18-JUN-1999;	99US-0139462.
PR	18-JUN-1999;	99US-0139463.
PR	18-JUN-1999;	99US-0139750.
PR	18-JUN-1999;	99US-0139763.
PR	21-JUN-1999;	99US-0139817.
PR	22-JUN-1999;	99US-0139899.
PR	23-JUN-1999;	99US-0140353.
PR	23-JUN-1999;	99US-0140354.
PR	24-JUN-1999;	99US-0140695.
PR	28-JUN-1999;	99US-0140823.
PR	29-JUN-1999;	99US-0140991.
PR	30-JUN-1999;	99US-0141287.
PR	01-JUL-1999;	99US-0141842.
PR	01-JUL-1999;	99US-0142154.
PR	02-JUL-1999;	99US-0142055.
PR	06-JUL-1999;	99US-0142390.
PR	08-JUL-1999;	99US-0142803.
PR	09-JUL-1999;	99US-0142920.
PR	12-JUL-1999;	99US-0142977.
PR	13-JUL-1999;	99US-0143542.
PR	14-JUL-1999;	99US-0143624.
PR	15-JUL-1999;	99US-0144005.
PR	16-JUL-1999;	99US-0144085.
PR	16-JUL-1999;	99US-0144086.
PR	19-JUL-1999;	99US-0144325.
PR	19-JUL-1999;	99US-0144331.
PR	19-JUL-1999;	99US-0144332.
PR	19-JUL-1999;	99US-0144333.
PR	19-JUL-1999;	99US-0144334.
PR	19-JUL-1999;	99US-0144335.
PR	20-JUL-1999;	99US-0144352.
PR	20-JUL-1999;	99US-0144632.
PR	20-JUL-1999;	99US-0144884.
PR	21-JUL-1999;	99US-0144814.
PR	21-JUL-1999;	99US-0145086.
PR	21-JUL-1999;	99US-0145088.
PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.
PR	22-JUL-1999;	99US-0145192.
PR	23-JUL-1999;	99US-0145145.
PR	23-JUL-1999;	99US-0145218.
PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.
PR	27-JUL-1999;	99US-0145913.
PR	27-JUL-1999;	99US-0145918.
PR	27-JUL-1999;	99US-0145919.
PR	28-JUL-1999;	99US-0145951.
PR	02-AUG-1999;	99US-0146388.
PR	02-AUG-1999;	99US-0146389.
PR	03-AUG-1999;	99US-0147038.
PR	04-AUG-1999;	99US-0147204.
PR	04-AUG-1999;	99US-0147302.
PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	16-AUG-1999;	99US-0148684.
PR	17-AUG-1999;	99US-0149368.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.

```

PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      80.5%; Score 33; DB 21; Length 279;
Best Local Similarity 88.9%; Pred. No. 67;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
Db 8 MARSLLLPL 16

RESULT 27
AAW74766
ID AAW74766 standard; Protein; 43 AA.
XX AC AAW74766;
XX DT 19-JAN-1999 (first entry)
XX DE Human secreted protein encoded by gene 36 clone HHPBD40.
XX KE Human; secreted protein; testis; tumour; foetal brain tissue;
KW fusion protein; cancer; central nervous system; seizure;
KW diagnosis; neurodegenerative disease.
XX OS Homo sapiens.
XX PN WO9839448-A2.
XX PD 11-SEP-1998.
XX PF 06-MAR-1998; 98WO-US04493.
XX PR 02-OCT-1997; 97US-0061060.
XX PR 07-MAR-1997; 97US-0038621.
XX PR 07-MAR-1997; 97US-0040161.
XX PR 07-MAR-1997; 97US-0040162.
XX PR 07-MAR-1997; 97US-0040163.
XX PR 07-MAR-1997; 97US-0040333.

PR 07-MAR-1997; 97US-0040334.
PR 07-MAR-1997; 97US-0040336.
PR 07-MAR-1997; 97US-0040626.
PR 11-APR-1997; 97US-0043311.
PR 11-APR-1997; 97US-0043312.
PR 11-APR-1997; 97US-0043313.
PR 11-APR-1997; 97US-0043314.
PR 11-APR-1997; 97US-0043568.
PR 11-APR-1997; 97US-0043569.
PR 11-APR-1997; 97US-0043576.
PR 11-APR-1997; 97US-0043578.
PR 11-APR-1997; 97US-0043580.
PR 11-APR-1997; 97US-0043669.
PR 11-APR-1997; 97US-0043670.
PR 11-APR-1997; 97US-0043671.
PR 11-APR-1997; 97US-0043672.
PR 11-APR-1997; 97US-0043674.
PR 23-MAY-1997; 97US-0047492.
PR 23-MAY-1997; 97US-0047500.
PR 23-MAY-1997; 97US-0047501.
PR 23-MAY-1997; 97US-0047502.
PR 23-MAY-1997; 97US-0047503.
PR 23-MAY-1997; 97US-0047581.
PR 23-MAY-1997; 97US-0047582.
PR 23-MAY-1997; 97US-0047583.
PR 23-MAY-1997; 97US-0047584.
PR 23-MAY-1997; 97US-0047585.
PR 23-MAY-1997; 97US-0047586.
PR 23-MAY-1997; 97US-0047587.
PR 23-MAY-1997; 97US-0047588.
PR 23-MAY-1997; 97US-0047589.
PR 23-MAY-1997; 97US-0047590.
PR 23-MAY-1997; 97US-0047592.
PR 23-MAY-1997; 97US-0047593.
PR 23-MAY-1997; 97US-0047594.
PR 23-MAY-1997; 97US-0047595.
PR 23-MAY-1997; 97US-0047596.
PR 23-MAY-1997; 97US-0047597.
PR 23-MAY-1997; 97US-0047598.
PR 23-MAY-1997; 97US-0047599.
PR 23-MAY-1997; 97US-0047600.
PR 23-MAY-1997; 97US-0047601.
PR 23-MAY-1997; 97US-0047612.
PR 23-MAY-1997; 97US-0047613.
PR 23-MAY-1997; 97US-0047614.
PR 23-MAY-1997; 97US-0047615.
PR 23-MAY-1997; 97US-0047617.
PR 23-MAY-1997; 97US-0047618.
PR 23-MAY-1997; 97US-0047632.
PR 23-MAY-1997; 97US-0047633.
PR 06-JUN-1997; 97US-0048964.
PR 06-JUN-1997; 97US-0048974.
PR 13-JUN-1997; 97US-0049610.
PR 08-JUL-1997; 97US-0051926.
PR 16-JUL-1997; 97US-0052874.
PR 18-AUG-1997; 97US-0055724.
PR 22-AUG-1997; 97US-0056630.
PR 22-AUG-1997; 97US-0056631.
PR 22-AUG-1997; 97US-0056632.
PR 22-AUG-1997; 97US-0056636.
PR 22-AUG-1997; 97US-0056637.
PR 22-AUG-1997; 97US-0056662.
PR 22-AUG-1997; 97US-0056664.
PR 22-AUG-1997; 97US-0056845.
PR 22-AUG-1997; 97US-0056862.
PR 22-AUG-1997; 97US-0056864.
PR 22-AUG-1997; 97US-0056872.
PR 22-AUG-1997; 97US-0056874.
PR 22-AUG-1997; 97US-0056875.
PR 22-AUG-1997; 97US-0056876.
PR 22-AUG-1997; 97US-0056877.
PR 22-AUG-1997; 97US-0056878.
PR 22-AUG-1997; 97US-0056879.

```

PR 22-AUG-1997; 97US-0056880.
 PR 22-AUG-1997; 97US-0056881.
 PR 22-AUG-1997; 97US-0056882.
 PR 22-AUG-1997; 97US-0056884.
 PR 22-AUG-1997; 97US-0056886.
 PR 22-AUG-1997; 97US-0056887.
 PR 22-AUG-1997; 97US-0056888.
 PR 22-AUG-1997; 97US-0056889.
 PR 22-AUG-1997; 97US-0056892.
 PR 22-AUG-1997; 97US-0056893.
 PR 22-AUG-1997; 97US-0056894.
 PR 22-AUG-1997; 97US-0056903.
 PR 22-AUG-1997; 97US-0056908.
 PR 22-AUG-1997; 97US-0056909.
 PR 22-AUG-1997; 97US-0056910.
 PR 22-AUG-1997; 97US-0056911.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057659.
 PR 05-SEP-1997; 97US-0057761.
 PR 12-SEP-1997; 97US-005785.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX

PI Bedharik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress GA;
 PI Peng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
 PI Kyaw H, Lafleur DW, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 XX

DR WPI; 1998-506364/43.

DR N-PSDB; AAV59546.

XX New isolated human genes and the secreted polypeptide(s) they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders

XX Claim 1; Page 561; 721pp; English.

XX This sequence represents a secreted human protein encoded by the nucleic
 CC acid molecule designated Gene 36 from the human cDNA clone HHPBD40
 CC (deposited as clone ATCC 97898 and ATCC 209044).

CC The gene can be used to generate fusion proteins by linking to the gene
 CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the
 CC stability of the fused protein as compared to the human protein only.

CC The invention relates to 186 novel genes and their fragments (nucleic
 CC acid sequences: AAV59511-V59812; amino acid sequences AAW4731-W75026)

CC which are useful for preventing, treating or ameliorating medical
 CC conditions e.g. by protein or gene therapy. Also, pathological

CC conditions can be diagnosed by determining the amount of the new
 CC polypeptides in a sample or by determining the presence of mutations in

CC the new polynucleotides. Specific uses are described for each of the 186
 CC polynucleotides, based on which tissues they are most highly expressed in
 CC (see AAV59511 for described uses).

XX Sequence 43 AA;
 SQ

Query Match 78.0%; Score 32; DB 19; Length 43;

Best Local Similarity 77.8%; Pred. No. 14;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9

1111111

Db 1 MKRSFLLPL 9

RESULT 28

ID AAB51724

XX AAB51724 standard; Protein; 54 AA.

AC AAB51724;

XX

DT 16-FEB-2001 (first entry)

XX

DE Human secreted protein sequence encoded by gene 1 SEQ ID NO:59.

XX

Human; secreted protein; immunosuppressive; antiarthritic; antirheumatic;
 KW antiproliferative; cytostatic; cardiant; vasotropic; cerebroprotective;
 KW nootropic; neuroprotective; antibacterial; virucide; fungicide;
 KW ophthalmological; vulnery; autoimmune disease; rheumatoid arthritis;
 KW hyperproliferative disorders; cancer; cardiovascular disorder;
 KW cardiac arrest; cerebrovascular disorder; nervous system disorder;
 KW Alzheimer's disease; ocular disorder; wound healing; skin aging.

XX Homo sapiens.

XX WO200061625-A1.

XX 19-OCT-2000.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

Query Match 78.0%; Score 32; DB 21; Length 54;

Best Local Similarity 100.0%; Pred. No. 18;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLLPL 9

1111111

Db 27 RSLLLPL 33

RESULT 29

AAU30194
 ID AAU30194 standard; Protein; 105 AA.
 XX
 AC AAU30194;
 XX
 DT 18-DEC-2001 (first entry)
 XX
 DE Novel human secreted protein #685.
 XX
 KW Human; vaccination; gene therapy; nutritional supplement;
 KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
 KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
 XX
 OS Homo sapiens.
 XX
 PN WO200179449-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 16-APR-2001; 2001WO-US08656.
 XX
 PR 18-APR-2000; 2000US-0552929.
 XX
 PR 26-JAN-2001; 2001US-0770160.
 XX
 PA (HYSEQ-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 XX
 DR WPI; 2001-611725/70.
 XX
 PT Nucleic acids encoding a range of human polypeptides, useful in genetic
 PT vaccination, testing and therapy -
 XX
 PS Claim 20: Page 258; 765pp; English.
 XX
 CC The invention relates to novel human secreted polypeptides. The
 CC polypeptides and antibodies to the polypeptides are useful for
 CC determining the presence of or predisposition to a disease associated
 CC with altered levels of polypeptide. The polypeptides are also useful for
 CC identifying agents (agonists and antagonists) that bind to them. Cells
 CC expressing the proteins are useful for identifying a therapeutic agent
 CC for use in treatment of a pathology related to aberrant expression or
 CC physiological interactions of the polypeptide. Vectors comprising
 CC the nucleic acids encoding the polypeptides and cells genetically
 CC engineered to express them are also useful for producing the proteins.
 CC The proteins are useful in genetic vaccination, testing and
 CC therapy, and can be used as nutritional supplements. They may be used to
 CC increase stem cell proliferation; to regulate haematopoiesis; and in
 CC bone, cartilage, tendon and/or nerve tissue growth or regeneration;
 CC immune suppression and/or stimulation; as anti-inflammatory agents; and
 CC in treatment of leukaemias. AAU29510-AAU33304 represent the amino acid
 CC sequences of novel human secreted proteins of the invention.
 XX
 SQ Sequence 105 AA;
 Query Match 78.0%; Score 32; DB 22; Length 105;
 Best Local Similarity 100.0%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 RSLLLPL 9
 Db 52 RSLLLPL 58
 RESULT 30
 ID AAB64397
 XX AAB64397 standard; Protein; 157 AA.
 AC AAB64397;
 XX
 XX 22-MAR-2001 (first entry)
 DT
 XX Amino acid sequence of human intracellular signalling molecule INTRA29.

XX
 KW Human; intracellular signalling molecule; INTRA; immunosuppressive;
 KW cytostatic; neuroprotective; nootropic; antiarteriosclerotic; cancer;
 KW antiinflammatory; anti-HIV; neuroleptic; antibacterial; antifungal;
 KW antiviral; antiparasitic; antihelminthic; antiparkinsonian; AIDS;
 KW cell proliferative disorder; arteriosclerosis autoimmune; epilepsy;
 KW inflammatory disorder; Addison's disease; gastrointestinal disorder;
 KW neurological disorder; Parkinson's disease; Creutzfeldt-Jakob disease;
 KW mental disorder; schizophrenia; anxiety.
 XX
 OS Homo sapiens.
 XX
 PN WO200077040-A2.
 XX
 PD 21-DEC-2000.
 XX
 PF 16-JUN-2000; 2000WO-US16636.
 XX
 PR 16-JUN-1999; 99US-0139566.
 XX
 PR 17-AUG-1999; 99US-0149640.
 XX
 PR 09-NOV-1999; 99US-0164417.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Yue H, Tang YT, Hillman JL, Lal P, Bandman O, Baughn MR;
 PI Azimzai Y, Yang J, Reddy R, Lu DAM;
 XX
 DR WPI; 2001-025334/03.
 XX
 DR N-PSDB; AAF32666.
 XX
 PT New human intracellular signalling molecules, useful for the diagnosis,
 PT prevention and treatment of cell proliferative, autoimmune,
 PT inflammatory, neurological, gastrointestinal, reproductive and
 PT developmental disorders -
 XX
 PS Claim 5: Page 138; 192pp; English.
 XX
 CC Sequences AAF32638 - AAF32689 represent cDNA encoding human
 CC intracellular signalling molecules INTRA - INTRA52, represented in
 CC AAB64369 - AAB64420. Modulators of the intracellular signalling molecules
 CC of the invention exhibit immunosuppressive; cytostatic; neuroprotective;
 CC nootropic; antiarteriosclerotic; antifungal; antiviral; antiparasitic;
 CC neuroleptic; antibacterial; antiparkinsonian activity. INTRA polypeptides their
 CC agonists and antagonists are useful for the treatment of a condition
 CC associated with decreased or increased expression of functional INTRA.
 CC Disorders associated with abnormal INTRA expression or activity include
 CC cell proliferative disorders e.g. arteriosclerosis and cancers;
 CC autoimmune or inflammatory disorders e.g. Addison's disease and acquired
 CC immunodeficiency syndrome (AIDS); viral, bacterial, fungal, parasitic,
 CC protozoal and helminthic infections; gastrointestinal disorders e.g.
 CC dysphagia and irritable bowel syndrome; neurological disorders e.g.
 CC epilepsy and Parkinson's disease; prion diseases e.g. Creutzfeldt-Jakob
 CC disease and mental disorders e.g. anxiety, schizophrenia and Tourette's
 CC disorder. Antibodies immuno specific for the INTRA proteins may also be
 CC useful in the diagnosis of the above disorders.
 XX
 SQ Sequence 157 AA;
 Query Match 78.0%; Score 32; DB 22; Length 157;
 Best Local Similarity 100.0%; Pred. No. 58;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 RSLLLPL 9
 Db 10 RSLLLPL 16
 RESULT 31
 ID AAB07446
 XX AAB07446 standard; Protein; 175 AA.
 AC AAB07446;

```

XX DT 20-OCT-2000 (first entry)
XX DE A human monocyte-derived protein FDF03-M14.
XX DE Human; monocyte-derived protein; FDF03; FDF03DeltaTM; FDF03-S1;
XX KW FDF03-M14; FDF03-S2; haematopoietic cell; monocyte hyperplasia;
XX KW tissue rejection; inflammation; infection.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Peptide 1..17
XX FT Protein /note= "signal sequence"
XX FT Protein 18..175
XX FT Protein /note= "mature protein"
XX PN WO200040721-A1.
XX PD 13-JUL-2000.
XX PF 29-DEC-1999; 99WO-US30004.
XX PR 31-DEC-1998; 98US-0223919.
XX PR 31-DEC-1998; 98US-0224604.
XX PA (SCHE ) SCHERING CORP.
XX PI Bates E, Fournier N, Chaulus L, Garrone P;
XX DR WPI: 2000-465984/40.
XX DR N-PSDB; AAA58817.
XX PT Novel monocyte-derived polypeptides and polynucleotides, used to
XX PT diagnose diseases associated with changes in monocyte numbers, e.g.
XX PT bacterial or viral infections -
XX PS Claim 1: Page 39-40; 45pp; English.
XX CC The present sequence represents a human monocyte-derived protein. The
XX CC specification describes monocyte-derived proteins FDF03, FDF03DeltaTM,
XX CC FDF03-S1, FDF03-M14, and FDF03-S2. The proteins are involved in the
XX CC regulation, or development, of haematopoietic cells. Antibodies specific
XX CC for antigenic components of the proteins can be used to detect the
XX CC components in samples. The proteins can also be used to screen for
XX CC candidate therapeutic agents. The monocyte-derived proteins and
XX CC polynucleotides can be used for diagnosis of diseases related to an
XX CC increase, or decrease, in the number of monocytes in a tissue or
XX CC lymph system, such as monocyte hyperplasia, tissue or graft rejection,
XX CC inflammation, or bacterial or viral infections. The proteins can also
XX CC be used in the treatment of disorders associated with abnormal expression
XX CC or signalling by a monocyte.
XX SQ Sequence 175 AA;

Query Match 78.0%; Score 32; DB 21; Length 175;
Best Local Similarity 77.8%; Pred. No. 66;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
Db 1 MGRPLLLPL 9

RESULT 32
AAW80407
ID AAW80407 standard; Protein; 226 AA.
XX AC AAW80407;
XX AC AAW80407;
XX DT 13-JAN-1999 (first entry)
XX DE A secreted protein encoded by clone di39_9.

```

```

XX KW Secreted protein; immune stimulating; suppressing;
XX KW haematopoiesis regulating activity; tissue growth activity; activin;
XX KW inhibin activity; chemotactic; chemokinetic activity; haemostatic;
XX KW thrombolytic activity; anti-inflammatory activity; cadherin;
XX KW tumour invasion suppressor activity; tumour inhibition activity.
XX OS Homo sapiens.
XX PN WO9844113-A1.
XX PD 08-OCT-1998.
XX PF 27-MAR-1998; 98WO-US06176.
XX PR 25-MAR-1998; 98US-0823330.
XX PR 28-MAR-1997; 97US-0823330.
XX PA (GEMY ) GENETICS INST INC.
XX PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
XX PI Racie LA, Spaulding V, Treacy M;
XX DR WPI: 1998-542703/46.
XX DR N-PSDB; AAV63191.
XX PT - New isolated polynucleotide(s) and secreted proteins - are obtained
XX PT from human cDNA libraries prepared from adult testes, foetal brain,
XX PT adult brain, adult blood and placenta
XX PS Claim 19; Pages 76-77; 124pp; English.
XX CC The present sequence represents a secreted protein. The nucleic acid
XX CC sequence is isolated from a human adult testes cDNA library using
XX CC probe AAV63202. The polypeptide may have biological activities such as
XX CC e.g. nutritional activity, immune stimulating or suppressing activity,
XX CC haematopoiesis regulating activity, tissue growth activity,
XX CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
XX CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory
XX CC activity, cadherin/tumour invasion suppressor activity, tumour
XX CC inhibition activity or other activities.
XX SQ Sequence 226 AA;

Query Match 78.0%; Score 32; DB 19; Length 226;
Best Local Similarity 77.8%; Pred. No. 87;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 MARSLLLPL 9
Db 1 MGRPLLLPL 9

RESULT 33
AAAY08015
ID AAAY08015 standard; Protein; 226 AA.
XX AC AAAY08015;
XX AC AAAY08015;
XX DT 08-JUL-1999 (first entry)
XX DE Human LSP-1 protein.
XX KW LSP-1; signal peptide; leucocyte-specific protein-1; PA-I; TAP-1; cancer;
XX KW proliferin analogue I; thrombopoietin analogue protein 1; anticancer;
XX KW antiangiogenic; anticancer; anti-inflammatory; anti-thrombocytopenic;
XX KW anti-arthritis; signal transduction; inflammatory; disease; growth;
XX KW proliferation; differentiation; cell survival; angiogenesis; diagnosis;
XX KW haematopoietic stem cell; erythroid precursor; megakaryocytopoiesis;
XX KW thrombopoiesis; prognosis; treatment; chromosome mapping; tissue typing;
XX KW forensics; arthritis; thrombocytopenia; bone marrow transplant; infection;
XX KW intravascular coagulation; iron deficiency; HIV; human.
XX

```

OS Homo sapiens.
 XX WO9918243-A1.
 PN
 XX 15-APR-1999.
 PD
 XX 06-OCT-1998; 98WO-US21151.
 PF
 XX 27-JAN-1998; 98US-0014347.
 PR 06-OCT-1997; 97US-0061143.
 PR 06-OCT-1997; 97US-0061149.
 PR 06-OCT-1997; 97US-0061159.
 PR 08-JAN-1998; 98US-0004206.
 PR 22-JAN-1998; 98US-0010674.
 XX
 PA (MILL-) MILLENNIUM BIOTHERAPEUTICS INC.
 XX
 XX Gearing DP, McCarthy SA, Pan Y;
 PI
 XX WPI; 1999-264042/22.
 DR N-PSDB; AAX37558.
 XX
 PT Signal-peptide containing proteins that modulate cellular processes
 XX
 PS Claim 2; Fig 1; 124pp; English.
 XX
 CC This invention describes the isolation of nucleic acids encoding the
 CC signal-peptide-containing molecules leucocyte-specific protein-1 (LSP-1),
 CC proliferin analog I (PA-I) and thrombopoietin analog protein 1 (TAP-1).
 CC These proteins have antiangiogenic, anticancer, anti-inflammatory, the
 CC anti-arthritis and anti-thrombocytopenic activity. The products of the
 CC invention and their modulators are involved in signal transduction,
 CC inflammatory responses, growth, proliferation, differentiation and
 CC survival of cells; angiogenesis; maturation of haematopoietic stem cells
 CC and erythroid precursors megakaryocytopoiesis and thrombopoiesis.
 CC Antibodies, or other binding agents, specific for the products of the
 CC invention are useful for diagnosis, prognosis and monitoring of treatment
 CC of diseases. Other uses include chromosome mapping, identification of
 CC individuals (tissue typing) and in forensic studies. LSP-1, PA-I and
 CC TAP-1 proteins and nucleic acids are modulators of cellular processes,
 CC particularly they are used to treat or prevent diseases associated with
 CC deregulation of angiogenesis, immune responses and haematopoiesis, e.g.
 CC cancer, arthritis (and other inflammatory diseases), thrombocytopenia
 CC (caused by cancer treatment, bone marrow transplant, human immune
 CC deficiency virus infection etc.), intravascular coagulation, iron
 CC deficiency etc.
 XX
 SQ Sequence 226 AA;
 Query Match 78.0%; Score 32; DB 20; Length 226;
 Best Local Similarity 77.8%; Pred. No. 87;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 MARSLLLPL 9
 I I I I I I I
 Db 1 MGRPLLLPL 9
 RESULT 34
 AAB07447
 ID AAB07447 standard; Protein; 226 AA.
 XX
 XX AAB07447;
 AC
 XX 20-OCT-2000 (first entry)
 DT
 XX
 XX A human monocyte-derived protein FDF03-S2.
 DE
 XX Human; monocyte-derived protein; FDF03; FDF03DeltaTM; FDF03-S1;
 KW FDF03-M14; FDF03-S2; haematopoietic cell; monocyte hyperplasia;
 KW tissue rejection; inflammation; infection.
 XX
 OS Homo sapiens.

XX Key Location/Qualifiers
 FH Peptide 1..17
 FT /note= "signal sequence"
 FT Protein 18..226
 FT /note= "mature protein"
 XX
 PN WO200040721-A1.
 XX
 PD 13-JUL-2000.
 XX
 PF 29-DEC-1999; 99WO-US30004.
 PR 31-DEC-1998; 98US-0223919.
 PR 31-DEC-1998; 98US-0224604.
 XX
 PA (SCHE) SCHERING CORP.
 XX
 PI Bates E, Fournier N, Chaulus L, Garrone P;
 XX
 DR WPI; 2000-465984/40.
 DR N-PSDB; AAA59818.
 XX
 PT Novel monocyte-derived polypeptides and polynucleotides, used to
 PT diagnose diseases associated with changes in monocyte numbers, e.g.
 PT bacterial or viral infections -
 XX
 PS Claim 1; Page 41-42; 45pp; English.
 XX
 CC The present sequence represents a human monocyte-derived protein. The
 CC specification describes monocyte-derived proteins FDF03, FDF03DeltaTM,
 CC FDF03-S1, FDF03-M14, and FDF03-S2. The proteins are involved in the
 CC regulation, or development, of haematopoietic cells. Antibodies specific
 CC for antigenic components of the proteins can be used to detect the
 CC components in samples. The proteins can also be used to screen for
 CC candidate therapeutic agents. The monocyte-derived proteins and
 CC polynucleotides can be used for diagnosis of diseases related to an
 CC increase, or decrease, in the number of monocytes in a tissue or
 CC lymph system, such as monocyte hyperplasia, tissue or graft rejection,
 CC inflammation, or bacterial or viral infections. The proteins can also
 CC be used in the treatment of disorders associated with abnormal expression
 CC or signalling by a monocyte.
 XX
 SQ Sequence 226 AA;
 Query Match 78.0%; Score 32; DB 21; Length 226;
 Best Local Similarity 77.8%; Pred. No. 87;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 MARSLLLPL 9
 I I I I I I I
 Db 1 MGRPLLLPL 9
 RESULT 35
 AAB07445
 ID AAB07445 standard; Protein; 227 AA.
 XX
 XX AAB07445;
 AC
 XX 20-OCT-2000 (first entry)
 DT
 XX
 XX A human monocyte-derived protein FDF03-S1.
 DE
 XX Human; monocyte-derived protein; FDF03; FDF03DeltaTM; FDF03-S1;
 KW FDF03-M14; FDF03-S2; haematopoietic cell; monocyte hyperplasia;
 KW tissue rejection; inflammation; infection.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..17
 FT /note= "signal sequence"
 FT

FT Protein 18..227
 TT /note= "mature protein"
 XX WO200040721-A1.
 PN 13-JUL-2000.
 XX 29-DEC-1999; 99WO-US30004.
 XX 31-DEC-1998; 98US-0223919.
 PR 31-DEC-1998; 98US-0224604.
 XX (SCHE) SCHERING CORP.
 XX Bates E, Fournier N, Chaulus L, Garrone P;
 XX WPI: 2000-465984/40.
 DR N-PSDB; AAA58816.
 XX Novel monocyte-derived polypeptides and polynucleotides, used to
 PT diagnose diseases associated with changes in monocyte numbers, e.g.
 PT bacterial or viral infections -
 XX Claim 1; Page 37-38; 45pp; English.
 XX The present sequence represents a human monocyte-derived protein. The
 CC specification describes monocyte-derived proteins FDF03, FDF03DeltaTM,
 CC FDF03-S1, FDF03-M14, and FDF03-S2. The proteins are involved in the
 CC regulation, or development, of haematopoietic cells. Antibodies specific
 CC for antigenic components of the proteins can be used to detect the
 CC components in samples. The proteins can also be used to screen for
 CC candidate therapeutic agents. The monocyte-derived proteins and
 CC polynucleotides can be used for diagnosis of diseases related to an
 CC increase, or decrease, in the number of monocytes in a tissue or
 CC lymph system, such as monocyte hyperplasia, tissue or graft rejection,
 CC inflammation, or bacterial or viral infections. The proteins can also
 CC be used in the treatment of disorders associated with abnormal expression
 CC or signalling by a monocyte.
 XX Sequence 227 AA;
 SQ
 Query Match 78.0%; Score 32; DB 21; Length 227;
 Best Local Similarity 77.8%; Pred. No. 87;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MARSLLLPL 9
 | | | | |
 Db 1 MGRPLLLPL 9
 RESULT 36
 AAY87230
 ID AAY87230 standard; Protein: 227 AA.
 XX AAY87230;
 AC 11-MAY-2000 (first entry)
 DT Human signal peptide containing protein HSPP-7 SEQ ID NO:7.
 DE Human; signal peptide-containing protein; HSPP; diagnosis; cancer;
 XX inflammation; cardiovascular disease; anticancer; anti-inflammatory;
 KW antimicrobial; neutropic; neuroprotective; cardiovascular; hepatotropic;
 KW antiasthmatic; gene therapy; cell proliferation; neurological disorder;
 KW reproductive disorder; developmental disorder; arteriosclerosis;
 KW cirrhosis; psoriasis; acquired immune deficiency syndrome; anaemia;
 KW asthma; Crohn's disease; infection; Alzheimer's disease; schizophrenia;
 KW Parkinson's disease; Huntington's diseases; ovulatory defect;
 XX muscular dystrophy.
 XX Homo sapiens.
 OS
 XX WO200000610-A2.
 PN

XX 06-JAN-2000.
 PD 25-JUN-1999; 99WO-US14484.
 XX 26-JUN-1998; 98US-0090762.
 PR 31-JUL-1998; 98US-0094983.
 PR 01-OCT-1998; 98US-0102686.
 PR 11-DEC-1998; 98US-0112129.
 XX (INCY-) INCYTE PHARM INC.
 PA Lal P, Tang YT, Gorgone GA, Corley NC, Guegler KJ, Baughn MR;
 PI Akerblom IE, Au-Young J, Yue H, Patterson C, Reddy R, Hillman JL;
 PI Bandman O;
 XX WPI: 2000-160673/14.
 DR N-PSDB; AA298115.
 XX New human signal peptide-containing proteins useful in treatment,
 PT prevention and diagnosis of e.g. cancer, inflammation and
 PT cardiovascular disease -
 XX Claim 1; Page 164-165; 327pp; English.
 XX AA298109 to AA298242 encode AAY87224 to AAY87357 which represent the
 CC human signal peptide-containing proteins HSPP-1 to HSPP-134. HSPPs have
 CC anticancer, anti-inflammatory, antimicrobial, neutropic, hepatotropic,
 CC neuroprotective, cardiovascular and antiasthmatic activities, and can
 CC be used in gene therapy. HSPPs can be used to treat or prevent disorders
 CC associated with decreased activity or function of HSPP. Antagonists of
 CC HSPP are used to treat or prevent disorders associated with increased
 CC activity or function of HSPP. Such diseases include cell proliferation
 CC (including cancer), inflammation, cardiovascular, neurological,
 CC reproductive or developmental disorders, (e.g. arteriosclerosis,
 CC cirrhosis, psoriasis, acquired immune deficiency syndrome, anaemia,
 CC asthma, Crohn's disease, microbial or other infections, congestive or
 CC ischaemic heart disease, Alzheimer's, Parkinson's or Huntington's
 CC diseases, schizophrenia, ovulatory defects, muscular dystrophy). HSPP
 CC nucleic acids can be used for the recombinant production of HSPP, for
 CC detecting HSPP in standard hybridisation and amplification assays (for
 CC diagnosis and monitoring), in gene therapy, as antisense,
 CC triplex-forming or ribozyme therapeutics, for detecting related sequences
 CC or genetic variations, and for chromosomal mapping. HSPP are also used to
 CC raise specific antibodies (Ab) and to screen for agonists and
 CC antagonists (potential therapeutic agents). Ab are used to diagnose, or
 CC monitor, HSPP-related diseases (in usual immunoassays), as therapeutic
 CC antagonists, in competitive drug screens, and for purification of HSPP
 CC from natural sources.
 XX Sequence 227 AA;
 SQ
 Query Match 78.0%; Score 32; DB 21; Length 227;
 Best Local Similarity 77.8%; Pred. No. 87;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MARSLLLPL 9
 | | | | |
 Db 1 MGRPLLLPL 9
 RESULT 37
 AAB07444
 ID AAB07444 standard; Protein: 230 AA.
 XX AAB07444;
 AC 20-OCT-2000 (first entry)
 XX A human monocyte-derived protein FDF03DeltaTM.
 DE Human; monocyte-derived protein; FDF03; FDF03DeltaTM; FDF03-S1;
 KW FDF03-M14; FDF03-S2; haematopoietic cell; monocyte hyperplasia;
 KW

KW tissue rejection; inflammation; infection.

OS Homo sapiens.

XX Key Location/Qualifiers

XX Peptide 1..17

FT Protein /note= "signal peptide"

FT 18..230

XX /note= "mature protein"

XX WO200040721-A1.

XX 13-JUL-2000.

XX 29-DEC-1999; 99WO-US30004.

XX 31-DEC-1998; 98US-0223919.

XX 31-DEC-1998; 98US-0224604.

XX (SCHE) SCHERING CORP.

XX Bates E, Fournier N, Chaulus L, Garrone P;

XX WPI: 2000-465984/40.

XX N-PSDB; AAA58815.

XX Novel monocytic-derived polypeptides and polynucleotides, used to
PT diagnose diseases associated with changes in monocyte numbers, e.g.
PT bacterial or viral infections -

XX Claim 1; Page 34-35; 45pp; English.

XX The present sequence represents a human monocyte-derived protein. The
CC specification describes monocyte-derived proteins FDF03, FDF03DeltaTM,
CC FDF03-S1, FDF03-M14, and FDF03-S2. The proteins are involved in the
CC regulation, or development, of haematopoietic cells. Antibodies specific
CC for antigenic components of the proteins can be used to detect the
CC components in samples. The proteins can also be used to screen for
CC candidate therapeutic agents. The monocyte-derived proteins and
CC polynucleotides can be used for diagnosis of diseases related to an
CC increase, or decrease, in the number of monocytes in a tissue or
CC lymph system, such as monocyte hyperplasia, tissue or graft rejection,
CC inflammation, or bacterial or viral infections. The proteins can also
CC be used in the treatment of disorders associated with abnormal expression
CC or signalling by a monocyte.

XX Sequence 230 AA;

Query Match 78.0%; Score 32; DB 21; Length 230;

Best Local Similarity 77.8%; Pred. No. 88;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9

Db 1 MGRPLLLPL 9

RESULT 38

ABG17241

ID ABG17241 standard; Protein; 251 AA.

XX

XX ABG17241;

XX

XX 18-FEB-2002 (first entry)

XX

XX Novel human diagnostic protein #17232.

XX

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

XX food supplement; medical imaging; diagnostic; genetic disorder.

XX

XX Homo sapiens.

XX

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB; AAS81428.

XX New isolated polynucleotide and encoded polypeptides, useful in

XX diagnostics, forensics, gene mapping, identification of mutations

XX responsible for genetic disorders or other traits and to assess

XX biodiversity -

XX

XX Claim 20; SEQ ID No 47600; 103pp; English.

XX

XX The invention relates to isolated polynucleotide (I) and

XX polypeptide (II) sequences. (I) is useful as hybridisation probes,

XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome

XX CC and gene mapping, and in recombinant production of (II). The

XX polynucleotides are also used in diagnostics as expressed sequence tags

XX for identifying expressed genes. (I) is useful in gene therapy techniques

XX to restore normal activity of (II) or to treat disease states involving

XX quantitating a polypeptide in tissue, as molecular weight markers and as

XX a food supplement. (II) and its binding partners are useful in medical

XX disorders involving aberrant protein expression or biological activity.

XX The polypeptide and polynucleotide sequences have applications in

XX diagnostics, forensics, gene mapping, identification of mutations

XX and to produce other types of data and products dependent on DNA and

XX amino acid sequences. ABG00010-ABG30377 represent novel human

XX diagnostic amino acid sequences of the invention.

XX Note: The sequence data for this patent did not appear in the printed

XX specification, but was obtained in electronic format directly from WIPO

XX at ftp.wipo.int/pub/published_pct_sequences.

XX

XX Sequence 251 AA;

XX

XX Query Match 78.0%; Score 32; DB 22; Length 251;

Best Local Similarity 77.8%; Pred. No. 97;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MARSLLLPL 9

Db 26 MAUSLLPL 34

RESULT 39

AAW63682

ID AAW63682 standard; Protein; 291 AA.

XX

XX AAW63682;

XX

XX 24-SEP-1998 (first entry)

XX

XX Human secreted protein 2.

XX

XX Secreted protein; human; cell proliferation; cytokine activity;

XX tissue growth; cellular differentiation; regeneration; activin;

XX inhibin; chemotactic; haemostatic; thrombolytic; tumour inhibition;

XX anti-inflammatory activity; biomarker.

XX

XX Homo sapiens.

XX

XX WO9825959-A2.

XX

XX

XX 18-JUN-1998.
PD
XX
PF 11-DEC-1997; 97WO-US22787.
XX
PR 11-DEC-1996; 96US-0032757.
PA (CHIR) CHIRON CORP.
PI Escobedo J, Garcia P, Hu Q, Kothakota S, Williams LT;
XX WPI; 1998-348453/30.
DR N-PSDB; AAV43602.
XX
XX Secreted human polypeptides - having cytokine, cell proliferation or
PT differentiation, activin or inhibin, tumour inhibition or
PT anti-inflammatory activities
XX
XX Claim 1; Pages 49-50; 78pp; English.
PS
XX This represents a human secreted protein. The specification provides
CC secreted protein sequences (AAW63681 to AAW63699) encoded by the nucleic
CC acid sequences shown in AAV43601 to AAV43619. The invention provides a
CC method of identifying a secreted polypeptide which is modified by rough
CC microsomes. The secreted proteins can be used in assays to determine
CC biological activities, such as cytokine, cell proliferation, or cellular
CC differentiation activities, tissue growth or regeneration, activin or
CC inhibin activity, chemotactic or chemokinetic activity, haemostatic or
CC thrombolytic activity, receptor/ligand activity, tumour inhibition, or
CC anti-inflammatory activity. The proteins can also be used as
CC biomarkers, to identify tissues or cell types which express the proteins,
CC or a stage- or disease-specific alteration in protein expression. They
CC can be used in protein interaction assays, to identify ligands or binding
CC proteins. Compounds which affect the biological activities of the
CC secreted proteins or their ability to interact with specific ligands can
CC be identified using the proteins in screening assays. The proteins and
CC antibodies that bind specifically to the protein can also be used to
CC design diagnostic tests and therapeutic compositions for diseases which
CC may be associated with altered expression of these proteins. Fusion
CC proteins comprising, e.g. signal sequences or transmembrane domains of
CC the proteins can be used to target other protein domains to cellular
CC membrane or they can be secreted extracellularly.
XX
XX Sequence 291 AA;
SQ
Query Match 78.0%; Score 32; DB 19; Length 291;
Best Local Similarity 77.8%; Pred. No. 1.1e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
Db 65 MGRPLLLPL 73
RESULT 40
AAW62772
ID AAW62772 standard; Protein; 303 AA.
XX
XX AAW62772;
AC
XX 23-SEP-1998 (first entry)
DT
DE Human immunoglobulin receptor designated FDF03.
XX
KW Human; type I transmembrane protein; immunoglobulin-like domain;
KW FDF03; activated monocyte; YE01; KTE03; control; development;
KW differentiation; mammalian immune system; treatment; cancerous condition;
KW degenerative condition; autoimmune response; transplantation rejection;
KW graft versus host disease; inflammatory condition; detection; diagnosis;
KW drug screening.
XX
XX Homo sapiens.
OS
XX

PN WO9824906-A2.
XX
PD 11-JUN-1998.
XX
PF 05-DEC-1997; 97WO-US21101.
XX
XX 21-MAR-1997; 97US-0041279.
PR 06-DEC-1996; 96US-0032252.
PR 09-DEC-1996; 96US-0762187.
PR 16-DEC-1996; 96US-0033181.
XX
XX (SCHE) SCHERING CORP.
PA
XX Adema GJ, Gorman DM, Lanier LL, McClanahan TK, Meyaard L;
PI Phillips JH, Zurawski G, Zurawski SM;
XX
XX WPI; 1998-333325/29.
DR N-PSDB; AAV38987.
XX
XX New isolated activated monocyte cell gene(s) - used to develop
PT products for treating e.g. cancer, degenerative conditions,
PT autoimmune responses, transplant rejection or inflammatory
PT conditions
XX
XX Claim 1; Pages 60-61; 104pp; English.
PS
XX The present sequence represents a human protein, FDF03, which is a type I
CC transmembrane protein comprising an extracellular portion characterised
CC by immunoglobulin-like domains, indicating that the protein is a
CC receptor member of the immunoglobulin superfamily. The FDF03 gene is
CC found in activated monocytes. The specification also describes other
CC proteins encoded by activated monocytes, which are designated YE01 and
CC KTE03. The genes function in controlling development, differentiation,
CC and/or physiology of the mammalian immune system. The products can be
CC used for treating abnormal proliferation, regeneration, degeneration or
CC atrophy. They can be used for treating e.g. cancerous conditions,
CC degenerative conditions, autoimmune responses, transplantation rejection,
CC graft versus host disease, or inflammatory conditions. The products can
CC also be used for detection, diagnosis and drug screening.
XX
XX Sequence 303 AA;
SQ
Query Match 78.0%; Score 32; DB 19; Length 303;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
Db 1 MGRPLLLPL 9
RESULT 41
AAB07443
ID AAB07443 standard; Protein; 303 AA.
XX
XX AAB07443;
AC
XX 20-OCT-2000 (first entry)
DT
DE A human monocyte-derived protein FDF03.
XX
KW Human; monocyte-derived protein; FDF03; FDF03deltaTM; FDF03-S1;
KW FDF03-M14; FDF03-S2; haematopoietic cell; monocyte hyperplasia;
KW tissue rejection; inflammation; infection.
XX
XX Homo sapiens.
OS
XX Key Location/Qualifiers
FT Peptide 1..19 /note= "signal peptide"
FT Protein 20..303 /note= "mature protein"
FT
XX

PN WO200040721-A1.
XX 13-JUL-2000.
XX 29-DEC-1999; 99WO-US30004.
XX 31-DEC-1998; 98US-0223919.
XX 31-DEC-1998; 98US-0224604.
XX (SCHE) SCHERING CORP.
XX
XX
PI Bates E, Fournier N, Chaulus L, Garrone P;
XX WPI; 2000-465984/40.
XX N-PSDB; AAA58814.
XX
XX Novel monocyte-derived polypeptides and polynucleotides, used to
XX diagnose diseases associated with changes in monocyte numbers, e.g.
XX bacterial or viral infections -
XX
XX Claim 1; Page 32-33; 45pp; English.
XX
XX The present sequence represents a human monocyte-derived protein. The
XX specification describes monocyte-derived proteins FDF03, FDF03DeltaTM,
XX FDF03-S1, FDF03-M14, and FDF03-S2. The proteins are involved in the
XX regulation, or development, of haematopoietic cells. Antibodies specific
XX for antigenic components of the proteins can be used to detect the
XX components in samples. The proteins can also be used to screen for
XX candidate therapeutic agents. The monocyte-derived proteins and
XX polynucleotides can be used for diagnosis of diseases related to an
XX increase, or decrease, in the number of monocytes in a tissue or
XX lymph system, such as monocyte hyperplasia, tissue or graft rejection,
XX inflammation, or bacterial or viral infections. The proteins can also
XX be used in the treatment of disorders associated with abnormal expression
XX or signalling by a monocyte.
XX
XX Sequence 303 AA;
XX
Query Match 78.0%; Score 32; DB 21; Length 303;
Best Local Similarity 77.8%; Pred. No. 1.2e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
DB 1 MGRPLLLPL 9
RESULT 42
ABBI2010
ID ABBI2010 standard; peptide; 326 AA.
XX
XX ABB12010;
XX
XX 11-JAN-2002 (first entry)
XX
XX Human secreted protein homologue, SEQ ID NO:2380.
XX
XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
XX haematopoiesis regulation; tissue growth; immunomodulator; activin;
XX inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
XX proliferation; metastasis; cancer; tumour; haematopoietic disorder;
XX myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
XX chronic inflammatory condition; proliferative retinopathy;
XX atherosclerosis; coronary heart disease; arterial ischaemia;
XX bone disorder; osteoporosis; vascular growth disorder;
XX tissue regeneration; wound healing; infection; immune disorder;
XX cell culture; drug screening; gene therapy; antiinflammatory;
XX antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
XX cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
XX antifungal; vulnery; antiulcer.
XX
XX Homo sapiens.
XX

PN WO200157188-A2.
XX 09-AUG-2001.
XX 05-FEB-2001; 2001WO-US03800.
XX 03-FEB-2000; 2000US-0496914.
XX 27-APR-2000; 2000US-0560875.
XX (HYSE*) HYSEQ INC.
XX
XX Tang YT, Liu C, Drmanac RT;
XX WPI; 2001-457740/49.
XX N-PSDB; ABA09254.
XX
XX Human proteins and DNA encoding sequences useful for preventing,
XX treating or ameliorating a medical condition in a mammalian subject
XX e.g. arthritis and cancer -
XX
XX Claim 20; Page 295; 1963pp; English.
XX
XX Sequences ABB10981-ABBI2330 represent 1350 novel human polypeptides, and
XX sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
XX invention also relates to vectors and recombinant host cells comprising a
XX nucleotide of the invention, methods of producing the novel polypeptides,
XX antibodies against the polypeptides, methods of detecting the nucleotides
XX or polypeptides in a sample, and methods of identifying compounds which
XX bind to polypeptides of the invention. Although novel, many of the
XX polypeptides of the invention have homology to known proteins, thereby
XX giving an insight into their probable biological activities, and hence
XX potential therapeutic applications. The polypeptides of the invention may
XX have various activities, including cytokine, cell proliferation or cell
XX differentiation activities; stem cell growth factor activity;
XX haematopoiesis regulatory activity; tissue growth activity;
XX immunomodulatory activity; activin- or inhibin-related activities;
XX chemotactic or chemokinetic activities; haemostatic, thrombotic or
XX thrombolytic activities; receptor or ligand activities; Or may be
XX involved in oncogenesis, cancer cell proliferation or metastasis.
XX Depending on their biological activities, polypeptides and nucleotides of
XX the invention are useful for preventing, treating or ameliorating medical
XX conditions, e.g., by protein or gene therapy. Such conditions include
XX cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
XX disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
XX proliferative retinopathy, atherosclerosis, coronary heart disease,
XX arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
XX vascular growth. Polypeptides involved with tissue regeneration and
XX repair (or nucleic acids encoding them) may be used to promote wound
XX healing (e.g., of burns, incisions and ulcers), while those with
XX immunomodulatory activities may be used in the treatment of viral,
XX bacterial and fungal infections in addition to immune disorders.
XX Polypeptides with growth factor activity may be used in cell cultures to
XX promote cell growth. For example, such polypeptides may be used to
XX manipulate stem cells in culture to give rise to neuroepithelial cells
XX that can be used to augment or replace cells damaged by illness,
XX autoimmune disease or accidental damage. The polypeptides and nucleotides
XX may also be used in the diagnosis of the above conditions, and in drug
XX screening techniques. The present sequence represents a novel human
XX polypeptide of the invention.
XX
XX Sequence 326 AA;
XX
Query Match 78.0%; Score 32; DB 22; Length 326;
Best Local Similarity 77.8%; Pred. No. 1.3e+02;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 MARSLLLPL 9
DB 100 MGRPLLLPL 108
RESULT 43
AAU29036

Db 2397 GPMVCRGGL 2405
||:||||| |

RESULT 7

Q9XSN6 PRELIMINARY; PRT; 254 AA.
AC Q9XSN6;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ENAMEL MATRIX SERINE PROTEINASE 1 PRECURSOR.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=98126310; PubMed=9465170;
RA Simmer J.P., Fukae M., Tanabe T., Yamakoshi Y., Uchida T., Xue J.,
RA Margolis H.C., Shimizu M., DeHart B.C., Hu C.-C., Bartlett J.D.;
RT "Purification, characterization, and cloning of enamel matrix serine
RT proteinase 1.",
RL J. Dent. Res. 77:377-386(1998).
CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; U76256; AAB94638.1; -;
DR HSSP; P00763; LDPO.
DR MEROPS; S01.251; -;
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 31 254 ENAMEL MATRIX SERINE PROTEINASE 1.
SQ SEQUENCE 254 AA; 27235 MW; FD40EF85664406F1 CRC64;

Query Match 80.0%; Score 40; DB 6; Length 254;
Best Local Similarity 66.7%; Pred. No. 4.7;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||:||||| |

Db 209 GPLICNGSL 217

RESULT 8

Q9XAC9 PRELIMINARY; PRT; 262 AA.
AC Q9XAC9;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE AMINOGLYCOSIDE ACETYLTRANSFERASE.
GN SCC22.09.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetales; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
SEQUENCE FROM N.A.
RX Seeger K., Harris D.;
RA Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rajandream M.A.;

Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

RL [3]
RN SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapate D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.",
RL MOL. Microbiol. 21:77-96(1996).
DR EMBL; AL096839; CAB50752.1; -;
DR InterPro: IPR003679; Antibiotic_NAT.
DR Pfam; PF02522; Antibiotic_NAT; 1.
KW Transferase.
SQ SEQUENCE 262 AA; 27930 MW; 0D239D41795A0B93 CRC64;

Query Match 80.0%; Score 40; DB 2; Length 262;
Best Local Similarity 88.9%; Pred. No. 4.8;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||||| |||| |

Db 8 GPLVTRGTL 16

RESULT 9

Q9YBC2 PRELIMINARY; PRT; 155 AA.
ID Q9YBC2
AC Q9YBC2;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 16.8 KDA PROTEIN APE1675.
GN APE1675.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.",
RL DNA Res. 6:83-101(1999).
DR EMBL; AP000062; BAA80676.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 155 AA; 16841 MW; 1AA47C685D52939B CRC64;

Query Match 78.0%; Score 39; DB 17; Length 155;
Best Local Similarity 77.8%; Pred. No. 4.6;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||:||||| |

Db 23 GPLITRGTL 31

RESULT 10

Q9W7Q5 PRELIMINARY; PRT; 247 AA.
ID Q9W7Q5
AC Q9W7Q5;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRYPSINOGEN 3.
OS Paralichthys olivaceus (Flounder).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Pleuronectiformes;
 OC Pleuronectoidel; Paralichthyidae; Paralichthys.
 OX NCBI_TaxID=8255;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PANCREAS;
 RA Suzuki T., Srivastava A.S., Kurokawa T.;
 RT "Japanese flounder mRNA for trypsinogen 3";
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL: AB029752; BAA82364.2; -.
 DR HSSP: P00763; IDPO.
 DR MEROPS: S01.151;
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP_SPC; 1.
 DR PROSITE: PS00240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 247 AA; 26948 MW; DC4B647179DDDD972 CRC64;
 Query Match 78.0%; Score 39; DB 13; Length 247;
 Best Local Similarity 77.8%; Pred. No. 7.1;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GPLVCRGTL 9
 ||||| |
 DB 202 GPLVNGEL 210
 RESULT 11
 ID Q9W174 PRELIMINARY; PRT; 288 AA.
 AC Q9W174;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE CG15873 PROTEIN.
 GN CG15873
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
 RA Kil H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., Li J., Li Z., Liang Y., Lin X.,
 RA Merkulov G., Milshina N.V., Mobarri C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY.
 DR EMBL: AE003464; AAF47203.1; -.
 DR HSSP: P20160; IA7S.
 DR MEROPS: S01.UNA;
 DR FlyBase: FBgn0035003; CG15873.
 DR InterPro: IPR001314; Chymotrypsin.
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin; 1.
 DR PRINTS: PR00722; CHYMOTRYPSIN.
 DR SMART: SM00020; TRYP_SPC; 1.
 DR PROSITE: PS00240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00134; TRYPSIN_HIS; 1.
 KW Hydrolase; Serine protease.
 SQ SEQUENCE 288 AA; 31961 MW; 36E5A4712516387A CRC64;
 Query Match 78.0%; Score 39; DB 5; Length 288;
 Best Local Similarity 66.7%; Pred. No. 8.2;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GPLVCRGTL 9
 ||||| |
 DB 231 GPLCKGAL 239
 RESULT 12
 ID Q9NR68 PRELIMINARY; PRT; 119 AA.
 AC Q9NR68;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE SERINE PROTEASE KALLIKREIN/OVASIN/NEUROPSIN TYPE 3.
 GN KLB8.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE=21206420; PubMed=11309326;
 RA Magklara A., Scorilas A., Katsaros D., Massobrio M., Yousef G.M.,
 RA Fracchioli S., Danese S., Diamandis E.P.;
 RA "The Human KLB8 (Neuropsin/Ovasin) Gene: Identification of Two Novel
 RT Splice Variants and Its Prognostic Value in Ovarian Cancer.";
 RL Clin. Cancer Res. 7:806-811(2001).
 DR EMBL: AF251125; AAF79144.1; -.
 DR HSSP: Q61955; INPM.
 DR MEROPS: S01.244;
 DR InterPro: IPR001254; Trypsin.
 DR Pfam: PF00089; trypsin; 1.
 DR SMART: SM00020; TRYP_SPC; 1.
 DR PROSITE: PS00240; TRYPSIN_DOM; 1.
 DR PROSITE: PS00135; TRYPSIN_SER; 1.

KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 119 AA; 12718 MW; 2FD8164DF1641EFF CRC64;

Query Match 76.0%; Score 38; DB 4; Length 119;
Best Local Similarity 77.8%; Pred. No. 5.6;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 73 GPLVCDGAL 81

RESULT 13
Q9CV76

ID Q9CV76 PRELIMINARY; PRT; 234 AA.
AC Q9CV76;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 2310008B01RIK PROTEIN (FRAGMENT).
GN 2310008B01RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC EMBL: AK009217; BAB26143.1; -.
DR HSSP: P00763; 1DPO.
DR MGD: MGI:1916761; 2310008B01RIK.
DR InterPro: IPR001314; Chymotrypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYPSIN_DOM; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 234 AA; 25888 MW; 6D81E609EDD39110 CRC64;

Query Match 76.0%; Score 38; DB 11; Length 234;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 188 GPLVCGVL 96

RESULT 14
Q9R048

ID Q9R048 PRELIMINARY; PRT; 234 AA.
AC Q9R048;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE STRATUM CORNEUM CHYMOTRYPTIC ENZYME (FRAGMENT).
GN K1K7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL;
RX MEDLINE=99399282; PubMed=10469296;
RA Backman A., Stranden P., Brattsand M., Hansson L., Egelrud T.,
RT "Molecular cloning and tissue expression of the murine analog to human
stratum corneum chymotryptic enzyme.";
RL J. Invest. Dermatol. 113:152-155(1999).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC EMBL: AF124299; AAF01139.1; -.
DR HSSP: P00757; 1SGF.
DR MEROPS: S01.300; -.
DR MGD: MGI:1346336; K1K7.
DR InterPro: IPR001314; Chymotrypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYPSIN_DOM; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00134; TRYPSIN_HIS; UNKNOWN_1.
KW Hydrolase; Serine protease.
FT NON_TER 234
SQ SEQUENCE 234 AA; 25220 MW; BEBE57D6C325B85F CRC64;

Query Match 76.0%; Score 38; DB 11; Length 234;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 203 GPLVCGDTL 211

RESULT 15
Q91VE3

ID Q91VE3 PRELIMINARY; PRT; 249 AA.
AC Q91VE3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE THYMOPSPIN (STRATUM CORNEUM CHYMOTRYPTIC ENZYME).
GN K1K7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THYMUS;
RA Yamaguchi N.;
RT "A novel cDNA cloning of mouse serine protease, Thymopsin, i.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RA Hansson L., Backman A., Ny A., Edlund M., Edholm E., Tornell J.,
RA Wallbrandt P., Egelrud T.;

RT "Epidermal overexpression of stratum corneum chymotryptic enzyme in
RL mice; a model for chronic itchy dermatitis.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB008371; BAB55604.1; -
DR EMBL: AF339930; AAK69852.1; -
SQ SEQUENCE 249 AA; 27257 MW; 0D4E380F12D14F97 CRC64;

Query Match 76.0%; Score 38; DB 11; Length 249;
Best Local Similarity 77.8%; Pred. No. 11;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPLVCRGTL 9
||||| ||
DB 203 GPLVCRGTL 211

RESULT 16

Q61096 PRELIMINARY; PRT; 254 AA.
ID Q61096; O08809;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-AUG-1999 (TrEMBLrel. 11, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MYELOBLASTIN PRECURSOR (EC 3.4.21.76) (PROTEINASE 3) (PROTEINASE-3)
DE (PR-3) (PRE-PRO-PROTEASE 3).
GN PRN3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=99126347; PubMed=9925946;
RA Sturrock A., Franklin K.F., Wu S.-Q., Hoidal J.R.;
RT "Characterization and localization of the genes for mouse proteinase-3
RT (Prtn3) and neutrophil elastase (Ela2).";
RL Cytogenet. Cell Genet. 83:104-108(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SV;
RX MEDLINE=97330929; PubMed=9187364;
RA Jenne D.E., Froehlich L., Hummel A.M., Specks U.;
RT "Cloning and functional expression of the murine homologue of
RT proteinase 3: implications for the design of murine models of
RT vasculitis.";
RL FEBS Lett. 408:187-190(1997).
RN [3]
RP SEQUENCE OF 2-254 FROM N.A.
RC STRAIN=BALB/C;
RX MEDLINE=97362044; PubMed=9211743;
RA Aveskogh M., Lutzelschwab C., Huang M.R., Hellman L.;
RT "Characterization of cDNA clones encoding mouse proteinase 3
RT (myeloblastin) and cathepsin G.";
RL Immunogenetics 46:181-191(1997).
CC -1- FUNCTION: POLYMORPHONUCLEAR LEUKOCYTE SERINE PROTEASE THAT
CC DEGRADES ELASTIN, FIBRONECTIN, LAMININ, VITRONECTIN, AND COLLAGEN
CC TYPES I, III, AND IV (IN VITRO) AND CAUSES EMPHYSEMA WHEN
CC ADMINISTERED BY TRACHEAL INSUFFLATION TO HAMSTERS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF PROTEINS, INCLUDING ELASTIN, BY
CC PREFERENTIAL CLEAVAGE: ALA-|-XAA > VAL-|-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN-FAMILY ELASTASE SUBFAMILY.
DR EMBL: AF082186; AAC79701.1; -
DR EMBL: U43525; AAB67271.1; -
DR EMBL: U97073; AAB58055.1; -
DR HSSP: P24158; 1FUJ.
DR MEROPS: S01.134; -
DR MGD: MGI:893580; Prtn3
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.

FT SIGNAL 1 27 BY SIMILARITY.
FT PROPEP 28 29 BY SIMILARITY.
FT CHAIN 30 250 MYELOBLASTIN.
FT PROPEP 251 254 BY SIMILARITY.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 127 127 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 176 176 N-LINKED (GLCNAC. .) (POTENTIAL).
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 154 211 BY SIMILARITY.
FT DISULFID 184 190 BY SIMILARITY.
FT DISULFID 201 226 BY SIMILARITY.
FT CONFLICT 2 2 S->A (IN REF. 2).
SQ SEQUENCE 254 AA; 27626 MW; 00CEB989A3CB79CA CRC64;

Query Match 76.0%; Score 38; DB 11; Length 254;
Best Local Similarity 66.7%; Pred. No. 11;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPLVCRGTL 9
||||| ||
DB 207 GPLVCRGTL 215

RESULT 17

Q9CS05 PRELIMINARY; PRT; 514 AA.
ID Q9CS05
AC Q9CS05; 2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 2510002A14RIK PROTEIN (FRAGMENT).
GN 2510002A14RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRYONIC LIVER;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J.,
RA Schram L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-P.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kotsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK010886; BAB27247.1; -
DR MGD: MGI:1919563; 2510002A14RIK.
FT NON_TER 1
SQ SEQUENCE 514 AA; 57637 MW; ED827FB86B2164B6 CRC64;

Query Match 76.0%; Score 38; DB 11; Length 514;

```

Best Local Similarity 87.5%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 PLVCRGRTL 9
    |||||
Db 98 PLVCRGIL 105

RESULT 18
Q62540 Q62540 PRELIMINARY; PRT; 46 AA.
AC Q62540;
DT 01-JUN-1998 (TrEMBLrel. 06, Created)
DT 01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GLANDULAR KALLIKREIN K5 (EC 3.4.21.35), SUBMANDIBULAR
DE (TISSUE KALLIKREIN) (MGK-5) (FRAGMENT).
GN KIK5.
OS Mus spretus (Western wild mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10096;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRET/ET; PubMed=8043949;
RX KO M.S., Wang X., Horton J.H., Hagen M.D., Takahashi N., Maezaki Y.,
RA Nadeau J.H.;
RT "Genetic mapping of 40 cDNA clones on the mouse genome by PCR.";
RL Mamm. Genome 5:349-355(1994).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE OF ARG-L-XAA BONDS IN
CC SMALL MOLECULE SUBSTRATES. HIGHLY SELECTIVE ACTION TO RELEASE
CC KALLIDIN (LYSYL-BRADYKININ) FROM KININOGEN INVOLVES HYDROLYSIS OF
CC MET-L-XAA OR LEU-L-XAA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. BELONGS TO THE GLANDULAR KALLIKREIN SUBFAMILY.
DR EMBL: U05716; AAB60472.1;
DR HSSP: P36368; 1A05
DR MGD: MGI:892020; Kik5.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
KW Hydrolyase; Serine protease; Kininogenase; Glycoprotein;
KW Multigene family; Zymogen.
FT NON_TER 1
FT NON_TER 46
FT NON_TER 46
SQ SEQUENCE 46 AA; 4938 MW; DC56A6D07B86580B CRC64;

Query Match 74.0%; Score 37; DB 11; Length 46;
Best Local Similarity 66.7%; Pred. No. 3.5;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGRTL 9
    |||:|
Db 1 GPLICDGVIL 9

RESULT 19
Q9CVU2 Q9CVU2 PRELIMINARY; PRT; 108 AA.
AC Q9CVU2;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE NERVE GROWTH FACTOR, GAMMA (FRAGMENT).
GN NGFG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]

```

```

SEQUENCE FROM N.A.
RP STRAIN=C57BL/6J; TISSUE=TESTIS;
RC MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mezzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK006516; BAB24629.1;
DR HSSP: P00756; ISGF.
DR MGD: MGI:97322; Ngfg.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00020; Tryp_SPC; 1.
DR PROSITE: PS50240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 108 AA; 11586 MW; 54DB958795AB9433 CRC64;

Query Match 74.0%; Score 37; DB 11; Length 108;
Best Local Similarity 66.7%; Pred. No. 7.9;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGRTL 9
    |||:|
Db 62 GPLICDGVIL 70

RESULT 20
Q9D974 Q9D974 PRELIMINARY; PRT; 114 AA.
AC Q9D974;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 1700127D06RIK PROTEIN.
GN 1700127D06RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TESTIS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,

```

```

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyoko-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK007298; BAB2494.1; -.
DR HSSP: P36368; IAO5.
DR MGD: MGI:1924249; 1700127D06Rik.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 114 AA; 13082 MW; CF6C05A967C47546 CRC64;

Query Match          74.0%; Score 37; DB 11; Length 114;
Best Local Similarity 66.7%; Pred. No. 8.4;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 58 GPLICDGL 66
|||||

RESULT 21
Q9PUF3
ID Q9PUF3 PRELIMINARY; PRT; 117 AA.
AC Q9PUF3;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRYPSINOGEN HOMOLOG (FRAGMENT).
OS Bothrops jararaca (Jararaca).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosaurs; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Bothrops.
OX NCBI_TaxID=8724;
RN [1]
SEQUENCE FROM N.A.
RA Murbach A.F., Hayashi M.A.F., Camargo A.C.M.;
RA "Screening of a Bothrops jararaca pancreas cDNA library.";
RT Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF190273; AAF01343.1; -.
DR HSSP: P00763; IDPO.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 117 AA; 12059 MW; 6894798659AB096C CRC64;

Query Match          74.0%; Score 37; DB 13; Length 117;
Best Local Similarity 66.7%; Pred. No. 8.6;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 73 GPVVCNAL 81
|||||

RESULT 22
Q9D6K1
ID Q9D6K1 PRELIMINARY; PRT; 128 AA.
AC Q9D6K1;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

```

```

DE 2900005P22RIK PROTEIN.
GN 2900005P22RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=HIPPOCAMPUS;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seva T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyoko-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK013485; BAB28877.1; -.
DR MGD: MGI:1920139; 2900005P22RIK.
SQ SEQUENCE 128 AA; 12955 MW; CE050940D6CABC24 CRC64;

Query Match          74.0%; Score 37; DB 11; Length 128;
Best Local Similarity 77.8%; Pred. No. 9.3;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 13 GPLVCPGL 21
|||||

RESULT 23
Q62284
ID Q62284 PRELIMINARY; PRT; 135 AA.
AC Q62284;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GAMMA-7S NERVE GROWTH FACTOR (Y-NGF) (FRAGMENT).
GN NGFG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=84169573; PubMed=6200835;
RA Howles P.N., Dickinson D.P., DiCaprio L.L., Woodworth-Gutai M.,
RA Gross K.W.;
RT "Use of a cDNA recombinant for the gamma subunit of mouse nerve growth
RT factor to localize members of this multigene family near the TAM-1
RT locus on chromosome 7.";
RL Nucleic Acids Res. 12:2791-2805(1984).
DR EMBL: X00472; CAA25154.1; -.
DR HSSP: P00756; ISGF.
DR MGD: MGI:97322; Ngfg.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00020; Tryp_Spc; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.

```

KW Hydrolase; Serine protease.

FT NON_TER 1
SQ SEQUENCE 135 AA; 14437 MW; E04055B826A92A48 CRC64;

Query Match 74.0%; Score 37; DB 11; Length 135;

Best Local Similarity 66.7%; Pred. No. 9.8;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

|||:| |

Db 89 GPLICDGL 97

RESULT 24

Q96JE1

ID Q96JE1 PRELIMINARY; PRT; 204 AA.

AC Q96JE1; 19, Created)

DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)

DE ARMI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PROSTATE;

RX MEDLINE=21398046; PubMed=11506707;

RA Korkmaz K.S., Korkmaz C.G., Pretlow T.G., Saatcioglu F.;

RT "Distinctly different gene structure of KLK4/KLK-L1/Protease/ARMI

RT compared with other members of the kallikrein family - Perinuclear

RT localization, alternative cDNA forms and regulation by multiple

RT hormones.";

RL DNA Cell Biol. 20:435-445(2001).

DR EMBL; AF259965; AAK71701.1; -.

SQ SEQUENCE 204 AA; 21811 MW; 45742165D23A53F3 CRC64;

Query Match

Best Local Similarity 74.0%; Score 37; DB 4; Length 204;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

|||:| |

Db 159 GPLICNGYL 167

RESULT 25

Q96JE2

ID Q96JE2 PRELIMINARY; PRT; 205 AA.

AC Q96JE2; 19, Created)

DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)

DE ARMI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PROSTATE;

RX MEDLINE=21398046; PubMed=11506707;

RA Korkmaz K.S., Korkmaz C.G., Pretlow T.G., Saatcioglu F.;

RT "Distinctly different gene structure of KLK4/KLK-L1/Protease/ARMI

RT compared with other members of the kallikrein family - Perinuclear

RT localization, alternative cDNA forms and regulation by multiple

RT hormones.";

RL DNA Cell Biol. 20:435-445(2001).

DR EMBL; AF259964; AAK71700.1; -.

SQ SEQUENCE 205 AA; 22052 MW; 159B61A20D94BD57 CRC64;

Query Match

Best Local Similarity 74.0%; Score 37; DB 4; Length 205;

Best Local Similarity 66.7%; Pred. No. 15;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

|||:| |

Db 160 GPLICNGYL 168

RESULT 26

Q96JE0

ID Q96JE0 PRELIMINARY; PRT; 205 AA.

AC Q96JE0;

DT 01-DEC-2001 (TEMBLrel. 19, Created)

DT 01-DEC-2001 (TEMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)

DE ARMI.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=PROSTATE TUMOR XENOGRAFT (CWR22);

RX MEDLINE=21398046; PubMed=11506707;

RA Korkmaz K.S., Korkmaz C.G., Pretlow T.G., Saatcioglu F.;

RT "Distinctly different gene structure of KLK4/KLK-L1/Protease/ARMI

RT compared with other members of the kallikrein family - Perinuclear

RT localization, alternative cDNA forms and regulation by multiple

RT hormones.";

RL DNA Cell Biol. 20:435-445(2001).

DR EMBL; AF259966; AAK71702.1; -.

SQ SEQUENCE 205 AA; 21950 MW; B83A025C73DBA1F6 CRC64;

Query Match

Best Local Similarity 74.0%; Score 37; DB 4; Length 205;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

|||:| |

Db 160 GPLICNGYL 168

RESULT 27

Q91036

ID Q91036 PRELIMINARY; PRT; 219 AA.

AC Q91036;

DT 01-NOV-1996 (TEMBLrel. 01, Created)

DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)

DE TRYPsinogen I (FRAGMENT).

OS Gadus morhua (Atlantic cod).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;

OC Gadus.

OX NCBI_TaxID=8049;

RN [1]

RP SEQUENCE FROM N.A.

RA Ong T.L., Armstrong R.F., McNamara P., Buckley L.J.;

RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

CC TRYPsin FAMILY.

DR EMBL; UA7819; AAB02196.1; -.

DR HSSP; P00763; IDPO.

DR InterPro; IPR001314; Chymotrypsin.

DR InterPro; IPR001254; Trypsin.

DR Pfam; PF00089; trypsin; 1.

DR PRINTS; PR00722; CHYMOTRYPSIN.

DR SMART; SM00020; Tryp.SPC; 1.

DR PROSITE; PS00240; TRYPsin_DOM; 1.

DR PROSITE; PS00134; TRYPsin_HIS; UNKNOWN_1.

DR PROSITE; PS00135; TRYPsin_SER; 1.

KW Hydrolase; Serine protease.

```
FT  NON_TER      1
SQ  SEQUENCE    219 AA;  23525 MW;  C96964EB49CEDIDA CRC64;
    Query Match      74.0%;  Score 37;  DB 13;  Length 219;
    Best Local Similarity 66.7%;  Pred. No. 16;
    Matches 6;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0;

QY  1 GPLVCRGTL 9
    DB 197 GPVVCNGVL 205
        ||:|:| |

RESULT 28
Q63274
ID  Q63274      PRELIMINARY;      PRT;      235 AA.
AC  Q63274;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  KALLIKREIN (FRAGMENT).
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=SPRAGUE-DAWLEY; TISSUE=KIDNEY;
RA  Zintz C.B., Ma J.-X., Chao J., Chao L.;
RT  "Isolation and characterization of a new rat kallikrein cDNA with
    predominant expression in the kidney.";
RL  Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
    TRYPSIN FAMILY.
CC  TRYPSIN FAMILY.
DR  EMBL; L33839; AAA58781.1; -.
DR  HSSP; P00759; ITON.
DR  MEROPS; S01.410; -.
DR  InterPro; IPR001314; Chymotrypsin.
DR  InterPro; IPR001254; Trypsin.
DR  Pfam; PF00089; trypsin; 1.
DR  PRINTS; PR00722; CHYMOTRYPSIN.
DR  SMART; SM00020; TRYSP-SPC; 1.
DR  PROSITE; PS00240; TRYPSIN_DOM; 1.
DR  PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR  PROSITE; PS00135; TRYPSIN_SER; 1.
KW  Hydrolase; Serine protease.
FT  NON_TER      1
SQ  SEQUENCE    235 AA;  26226 MW;  26160E9D3F7DC595 CRC64;

    Query Match      74.0%;  Score 37;  DB 11;  Length 235;
    Best Local Similarity 66.7%;  Pred. No. 17;
    Matches 6;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0;

QY  1 GPLVCRGTL 9
    DB 189 GPLICDGLV 197
        ||:|:| |

RESULT 29
Q63275
ID  Q63275      PRELIMINARY;      PRT;      239 AA.
AC  Q63275;
DT  01-NOV-1996 (TrEMBLrel. 01, Created)
DT  01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  KALLIKREIN (FRAGMENT).
OS  Rattus norvegicus (Rat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX  NCBI_TaxID=10116;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN=SPRAGUE-DAWLEY; TISSUE=SUBMANDIBULAR GLAND;
RA  Zintz C.B., Ma J.-X., Chao J., Chao L.;
```

```
RT  "Isolation and characterization of a new rat kallikrein cDNA with
    predominant expression in the kidney.";
RL  Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
    TRYPSIN FAMILY.
CC  TRYPSIN FAMILY.
DR  EMBL; L33840; AAA58782.1; -.
DR  HSSP; P00759; ITON.
DR  MEROPS; S01.411; -.
DR  InterPro; IPR001314; Chymotrypsin.
DR  InterPro; IPR001254; Trypsin.
DR  Pfam; PF00089; trypsin; 1.
DR  PRINTS; PR00722; CHYMOTRYPSIN.
DR  SMART; SM00020; TRYSP-SPC; 1.
DR  PROSITE; PS00240; TRYPSIN_DOM; 1.
DR  PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR  PROSITE; PS00135; TRYPSIN_SER; 1.
KW  Hydrolase; Serine protease.
FT  NON_TER      1
SQ  SEQUENCE    239 AA;  26382 MW;  3CBD761AF06ABB53 CRC64;

    Query Match      74.0%;  Score 37;  DB 11;  Length 239;
    Best Local Similarity 66.7%;  Pred. No. 17;
    Matches 6;  Conservative 1;  Mismatches 2;  Indels 0;  Gaps 0;

QY  1 GPLVCRGTL 9
    DB 193 GPLICDGLV 201
        ||:|:| |

RESULT 30
Q63301
ID  Q63301      PRELIMINARY;      PRT;      246 AA.
AC  Q63301;
DT  01-NOV-1998 (TrEMBLrel. 08, Created)
DT  01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT  01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE  SERINE PROTEASE (BSP).
GN  PRSS18 OR MBSP.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX  NCBI_TaxID=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=BRAIN;
RA  Matsui H., Takahashi T.;
RT  "Mouse serine protease preferentially expressed in brain.";
RL  Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC  -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
    TRYPSIN FAMILY.
CC  TRYPSIN FAMILY.
DR  EMBL; AB015206; BAA28895.1; -.
DR  EMBL; Y18723; CAA77269.1; -.
DR  EMBL; AB032402; BAA84544.1; -.
DR  HSSP; P00763; LDPO.
DR  MEROPS; S01.249; -.
DR  MGD; MGI:1343166; prss18.
DR  InterPro; IPR001314; Chymotrypsin.
DR  InterPro; IPR001254; Trypsin.
DR  Pfam; PF00089; trypsin; 1.
DR  PRINTS; PR00722; CHYMOTRYPSIN.
DR  SMART; SM00020; Tryp_SPC; 1.
```

```
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Protease; Serine protease.
SQ SEQUENCE 246 AA; 27496 MW; E20C080087139B63 CRC64;

Query Match 74.0%; Score 37; DB 11; Length 246;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGRTL 9
   |||||
DB 199 GPLVCGGRL 204

RESULT 31
O18458 PRELIMINARY; PRT; 247 AA.
AC O18458;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE SERINE PROTEINASE PRECURSOR.
GN SP-II.
OS Heterodera glycines (Soybean cyst nematode).
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
OC Tylenchoidea; Heteroderidae; Heteroderinae; Heterodera.
OX NCBI_TaxID=51029;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98030247; PubMed=9364965;
RA Lilley C.J., Urwin P.E., Atkinson H.J., McPherson M.J.;
RT "Characterization of cDNAs encoding serine proteinases from the
  soybean cyst nematode Heterodera glycines.";
RL Mol. Biochem. Parasitol. 89:195-207(1997).
DR EMBL; Y13907; CAA74205.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.09A; -.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 22 247 SERINE PROTEINASE.
SQ SEQUENCE 247 AA; 25586 MW; 3A2B5B2B3BB77222 CRC64;

Query Match 74.0%; Score 37; DB 5; Length 247;
Best Local Similarity 77.8%; Pred. No. 17;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGRTL 9
   |||||
DB 207 GPLVCGGRTL 215

RESULT 32
Q03955 PRELIMINARY; PRT; 250 AA.
AC Q03955;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE KALLIKREIN.
OS Praomys natalensis (African soft-furred rat) (Mastomys natalensis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
OC Mastomys.
OX NCBI_TaxID=10112;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RC TISSUE-SALIVARY GLAND;
RX MEDLINE=94226702; PubMed=7909667;
RA Fahnestock M.;
RT "Characterization of kallikrein cDNAs from the African rodent
  Mastomys.";
RL DNA Cell Biol. 13:293-300(1994).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
  TRYPSIN FAMILY.
DR EMBL; X17351; CAA35231.1; -.
DR HSSP; P00737; LSGF.
DR MEROPS; S01.160; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 250 AA; 27385 MW; 0D5908A2983C192F CRC64;

Query Match 74.0%; Score 37; DB 11; Length 250;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGRTL 9
   |||||
DB 213 GPLICDGLV 221

RESULT 33
Q91Y82 PRELIMINARY; PRT; 253 AA.
AC Q91Y82;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE NEUROGIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RA Mitsui S., Yamaguchi N.;
RT "cDNA cloning and characterization of mouse brain specific serine
  protease, Neurosin.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008928; BAB55605.1; -.
SQ SEQUENCE 253 AA; 28329 MW; C5EF98C7EEF2FBC1 CRC64;

Query Match 74.0%; Score 37; DB 11; Length 253;
Best Local Similarity 77.8%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGRTL 9
   |||||
DB 206 GPLVCGGRL 214

RESULT 34
Q96RQ0 PRELIMINARY; PRT; 255 AA.
AC Q96RQ0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE PROTIINGEN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```

OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21226193; PubMed=11327827;
RA Takayama T.K., Carter C.A., Deng T.;
RT "Activation of prostate-specific antigen precursor (pro-PSA) by
RT prostin, a novel human prostatic serine protease identified by
RT degenerate PCR.";
RL Biochemistry 40:1679-1687(2001).
DR EMBL: AF303046; AA62813.1; -.
SQ SEQUENCE 255 AA; 27986 MW; 00D5B79E14B9468F CRC64;

Query Match 74.0%; Score 37; DB 4; Length 255;
Best Local Similarity 77.8%; Pred. No. 18;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
   |||||
Db 210 GPLVCGGIL 218

RESULT 35
Q9R0K0 PRELIMINARY; PRT; 256 AA.
AC Q9R0K0;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GRANZYME K.
GN GZMK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=DDY;
MEDLINE=9937786; PubMed=10407175;
RA Suemoto T., Taniguchi M., Shiosaka S., Yoshida S.;
RT "CDNA cloning and expression of a novel serine protease in the mouse
RT brain.";
RL Brain Res. Mol. Brain Res. 70:273-281(1999).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL: AB032200; BAA84221.1; -.
DR HSSP; P00763; IDPO.
DR MEROPS; S01.146; -.
DR MGD; MGI:1298232; Gzmk.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT SIGNAL 1 24
FT CHAIN 25 261
SQ SEQUENCE 256 AA; 28586 MW; AC0CE7418EFD16FB CRC64;

Query Match 74.0%; Score 37; DB 11; Length 256;
Best Local Similarity 71.4%; Pred. No. 18;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
   |||:|
Db 208 GPLCKG 214

RESULT 36
Q29474 PRELIMINARY; PRT; 261 AA.
ID Q29474
AC Q29474;
DT 01-NOV-1996 (Tremblrel. 01, Created)

```

```

DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE KALLIKREIN PRECURSOR (EC 3.4.21.35).
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=MONGREL;
MEDLINE=94250683; PubMed=8193155;
RA Gauthier E.R., Dumas C., Chapdelaine P., Tremblay R.R., Dube J.Y.;
RT "Characterization of canine pancreas kallikrein cDNA.";
RL Biochim. Biophys. Acta 1218:102-104(1994).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; X75479; CAA53210.1; -.
DR HSSP; P00752; 2PKA.
DR MEROPS; S01.160; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL 1 24
FT CHAIN 25 261
SQ SEQUENCE 261 AA; 28913 MW; B5A49E1B66F34631 CRC64;

Query Match 74.0%; Score 37; DB 6; Length 261;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
   |||:|
Db 215 GPLICDGV 223

RESULT 37
Q9JW70 PRELIMINARY; PRT; 261 AA.
ID Q9JW70
AC Q9JW70;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE GLANDULAR KALLIKREIN 21 (SIMILAR TO KALLIKREIN 21).
GN KLK21 OR GK21.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20534760; PubMed=11082197;
RA Matsui H., Moriyama A., Takahashi T.;
RT "Cloning and characterization of mouse Klk27, a novel tissue
RT kallikrein expressed in testicular Leydig cells and exhibiting
RT chymotrypsin-like specificity.";
RL Eur. J. Biochem. 267:6858-6865(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX TISSUE-SALIVARY GLAND, 10 WEEK OLD FEMALE MOUSE;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AB039276; BAA92319.1; -.
DR EMBL; BC012243; AAH12243.1; -.
DR HSSP; P00757; 1SGF.
DR MEROPS; S01.038; -.

```

```
DR MGD; MGI:892022; Kik21.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS0135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 261 AA; 28690 MW; 608B976BC78E03EE CRC64;

Query Match 74.0%; Score 37; DB 11; Length 261;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 215 GPLICDGLV 223
|||||

RESULT 38
Q9JW71 Q9JW71 PRELIMINARY; PRT; 263 AA.
AC Q9JW71;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE GLANDULAR KALLIKREIN 27.
GN KUK27 OR GK27.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=20534760; PubMed=11082197;
RA Matsui H., Moriyama A., Takahashi T.;
RT "Cloning and characterization of mouse Kik27, a novel tissue
RT kallikrein expressed in testicular Leydig cells and exhibiting
RT chymotrypsin-like specificity.";
RL Eur. J. Biochem. 267:6858-6865(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR ENBL; AB039275; BAA92318.1; -.
DR HSP; P00757; 1SGF.
DR MEROPS; S01.073; -.
DR MGD; MGI:891980; Kik27.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS0240; TRYPSIN_DOM; 1.
DR PROSITE; PS0134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS0135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 263 AA; 28742 MW; 2B396C2DD3B359FB CRC64;

Query Match 74.0%; Score 37; DB 11; Length 263;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 217 GPLICDGLV 225
|||||

RESULT 39
Q9AQS5 Q9AQS5 PRELIMINARY; PRT; 311 AA.
AC Q9AQS5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
```

```
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CATECHOL 1,2-DIOXYGENASE (EC 1.13.11.1).
GN CATA.
OS Burkholderia sp. NK8.
OC Bacteria; Proteobacteria; beta subdivision; Burkholderia group;
OC Burkholderia.
OX NCBI_TaxID=140098;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NK8;
RX MEDLINE=21097249; PubMed=11160806;
RA Francisco P.B., Ogawa N., Suzuki K., Miyashita K.;
RT "The chlorobenzoate dioxygenase genes of Burkholderia sp. strain NK8
RT involved in the catabolism of chlorobenzoates.";
RL Microbiology 147:121-133(2001).
DR ENBL; AB024746; BAB21462.1; -.
DR HSP; P00437; 3PC.
DR InterPro; IPR000627; Dioxygenase.
DR Pfam; PF00775; Dioxygenase; 1.
DR PROSITE; PS00083; INTRADIOL_DIOXYGENAS; UNKNOWN_1.
KW Dioxygenase; Oxidoreductase.
SQ SEQUENCE 311 AA; 33773 MW; E828042A68E1DC09 CRC64;

Query Match 74.0%; Score 37; DB 2; Length 311;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 132 GPLVIRGTV 140
|||||

RESULT 40
Q9STY4 Q9STY4 PRELIMINARY; PRT; 599 AA.
AC Q9STY4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 65.5 KDA PROTEIN.
GN T21L8.140.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsals.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Choisne N., Robert C., Brottier P., Wincker P., Cattolico L.,
RA Artiguenave F., Saurin W., Weissenbach J., Mewes H.W., Mayer K.F.X.,
RA Lemcke K., Schueller C., Quetier F., Salanoubat M.;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR ENBL; AL096860; CAB51211.1; -.
DR InterPro; IPR002125; dCMP_cyt_deam.
DR InterPro; IPR002734; RibD_C.
DR Pfam; PF00383; dCMP_cyt_deam; 1.
DR Pfam; PF01872; RibD_C; 1.
KW Hypothetical protein.
SQ SEQUENCE 599 AA; 65502 MW; 11829D0332046797 CRC64;

Query Match 74.0%; Score 37; DB 10; Length 599;
Best Local Similarity 75.0%; Pred. No. 40;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PLVCRGTL 9
DB 12 PLICRATL 19
|||||
```



```

RESULT 41
Q61752
ID Q61752 PRELIMINARY; PRT; 66 AA.
AC Q61752;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE KALLIKREIN (FRAGMENT).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83182015; PubMed=6926406;
RA Nordeen S.K., Mason A.J., Richards R.I., Baxter J.D., Shine J.;
RT "Mouse kallikrein arginyl-esteropeptidase genes: Analysis of cloned
RT cDNAs suggests rapid functional divergence from a common ancestral
RT sequence."
RL DNA 1:309-311(1982).
DR EMBL: K01654; AAA39346.1; -.
DR HSSP: P36368; IA05.
DR MEROPS: S01.070; -.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
SQ SEQUENCE 66 AA; 6860 MW; DF2AAB0D012F0276 CRC64;

Query Match 72.0%; Score 36; DB 11; Length 66;
Best Local Similarity 66.7%; Pred. No. 7.8;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 20 GPLICDGIL 28

RESULT 42
Q9D725
ID Q9D725 PRELIMINARY; PRT; 138 AA.
AC Q9D725;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE 0910001B19RIK PROTEIN.
GN 0910001B19RIK.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=STOMACH;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,

```

```

RA Suzuki H., Toyo-oka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK008660; BAB25816.1; -.
DR HSSP: P00763; 1SLU.
DR MGD: MGI:1913350; 0910001B19RIK.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR SMART: SM00020; tryp_Spc; 1.
DR PROSITE: PS0240; TRYPSIN_DOM; 1.
DR PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
SQ SEQUENCE 138 AA; 14325 MW; 860D67A37BB5DEFA CRC64;

Query Match 72.0%; Score 36; DB 11; Length 138;
Best Local Similarity 66.7%; Pred. No. 16;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 94 GPVVCNGQL 102

RESULT 43
Q925V8
ID Q925V8 PRELIMINARY; PRT; 149 AA.
AC Q925V8;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MOUSE GLANDULAR KALLIKREIN MRNA, 3' END (FRAGMENT).
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82142394; PubMed=6174512;
RA Richards R.I., Cantanzaro D.F., Mason A.J., Morris B.J., Baxter J.D.,
RA Shine J.;
RT "Mouse glandular kallikrein genes. Nucleotide sequence of cloned cDNA
RT coding for a member of the kallikrein arginyl esteropeptidase group of
RT serine proteases.";
RL J. Biol. Chem. 257:2758-2761(1982).
DR EMBL: J00389; AAA37694.1; -.
FT NON_TER 1
SQ SEQUENCE 149 AA; 15972 MW; 1760957DF7383BD8 CRC64;

Query Match 72.0%; Score 36; DB 11; Length 149;
Best Local Similarity 66.7%; Pred. No. 17;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 103 GPLICDGIL 111

RESULT 44
Q93594
ID Q93594 PRELIMINARY; PRT; 178 AA.
AC Q93594;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TRYPSIN (EC 3.4.21.4) (FRAGMENT).
OS Dicotylarchus labrax (European sea bass).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Perciformes; Percoidae;
OC Moronidae; Dicentrarchus.
OX NCBI_TaxID=13489;

```

```

RN [1]
RA SEQUENCE FROM N.A.
RP Peres A., Zambonino Infante J.L., Cahu C.L.:
RT "Dietary regulation of activities and mRNA levels of trypsin and
RT amylase in sea bass (Dicentrarchus labrax) larvae.";
RL Fish Physiol. Biochem. 19:145-152(1998).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AJ006882; CAA07315.1; -.
DR HSSP; P35031; 1BIT.
DR MEROPS; S01.151; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS02440; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
FT NON_TER 178
FT NON_TER 178
SQ SEQUENCE 178 AA; 19352 MW; 13F13BEC80EDC57 CRC64;

Query Match 72.0%; Score 36; DB 13; Length 178;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 164 GPVVCNGEL 172
||:| | |
| | | | |

RESULT 45
Q9PT51 PRELIMINARY; PRT; 233 AA.
AC Q9PT51;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE BETA-FIBRINOGENASE (FRAGMENT).
GN BREVINASE.
OS Agkistrodon halys blomhoffi (Mamushi) (Gloydius blomhoffii).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Agkistrodon.
OX NCBI_TaxID=61300;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20302322; PubMed=10845699;
RA Lee J.W., Park W.;
RT "cDNA cloning of brevinase, a heterogeneous two-chain fibrinolytic
RT enzyme from Agkistrodon blomhoffii brevicaudus snake venom, by serial
RT hybridization-polymerase chain reaction.";
RL Arch. Biochem. Biophys. 377:234-240(2000).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
DR EMBL; AJ243757; CAB65936.1; -.
DR HSSP; P00763; 1DPO.
DR MEROPS; S01.185; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS02440; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 1
FT NON_TER 1
SQ SEQUENCE 233 AA; 25725 MW; 1676DC5AF0AB5A42 CRC64;

Query Match 72.0%; Score 36; DB 13; Length 233;

```

```

Best Local Similarity 55.6%; Pred. No. 26;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 181 GPLICNGEI 189
|||:| | |
| | | | |

Search completed: November 6, 2002, 12:12:11
Job time : 20.444 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 5.33333 Seconds
(without alignments)
65.339 Million cell updates/sec

Title: US-09-905-083-80
Perfect score: 50
Sequence: 1 GPLVCRGTL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	253	1 KLK7_HUMAN	P49862 homo sapien
2	45	90.0	250	1 KLK9_HUMAN	O9ukg9 homo sapien
3	44	88.0	251	1 KLKE_HUMAN	O9p0g3 homo sapien
4	40	80.0	258	1 VSP3_BOTJA	O9ptu8 bothrops ja
5	40	80.0	260	1 NRPN_RAT	O88780 rattus norv
6	40	80.0	261	1 KLK2_HUMAN	P20151 homo sapien
7	40	80.0	261	1 KLK3_HUMAN	P07288 homo sapien
8	39	78.0	261	1 KLK1_RAT	P00758 rattus norv
9	38	76.0	248	1 KLK2_HUMAN	O9ukr0 homo sapien
10	38	76.0	253	1 CFAD_HUMAN	P00746 homo sapien
11	38	76.0	258	1 VSP3_TRIGA	O13063 trimeresuru
12	38	76.0	260	1 KLK8_HUMAN	O60259 homo sapien
13	38	76.0	261	1 KLK3_MACMU	P33619 macaca mula
14	38	76.0	293	1 KLK5_HUMAN	O9y337 homo sapien
15	38	76.0	1154	1 CTA3_HUMAN	O9bz76 homo sapien
16	38	76.0	1308	1 CTA4_HUMAN	O9c0a0 homo sapien
17	38	76.0	1310	1 CTA4_MOUSE	O99b47 mus musculus
18	37	74.0	188	1 KLK3_RAT	P15950 rattus norv
19	37	74.0	209	1 CUT1_ALTR	P41744 alternaria
20	37	74.0	239	1 KLK2_CAVPO	P12323 cavia porce
21	37	74.0	241	1 TRY1_GADMO	P16049 gadus morhu
22	37	74.0	241	1 TRYX_GADMO	O91041 gadus morhu
23	37	74.0	244	1 KLK4_RAT	P36375 rattus norv
24	37	74.0	254	1 CTRL_HALRU	P35003 halloetis ru
25	37	74.0	256	1 KLK4_HUMAN	O9y5k2 homo sapien
26	37	74.0	256	1 KLK4_HUMAN	O9h2r5 homo sapien
27	37	74.0	258	1 GRK_RAT	P49864 rattus norv
28	37	74.0	259	1 KLK2_RAT	P00759 rattus norv
29	37	74.0	259	1 KLK9_RAT	P07647 rattus norv
30	37	74.0	259	1 KLK4_MOUSE	P15948 mus musculus
31	37	74.0	260	1 NRPN_MOUSE	O61955 mus musculus
32	37	74.0	261	1 KLK1_MOUSE	P15947 mus musculus
33	37	74.0	261	1 KLK3_MOUSE	P00756 mus musculus

34	37	74.0	261	1 KLK5_MOUSE	P15945 mus musculus
35	37	74.0	261	1 KLK6_MOUSE	P00755 mus musculus
36	37	74.0	261	1 KLK8_RAT	P36374 rattus norv
37	37	74.0	261	1 KLK9_MOUSE	P15949 mus musculus
38	37	74.0	261	1 KLKB_MOUSE	P15946 mus musculus
39	37	74.0	261	1 KLKG_MOUSE	P04071 mus musculus
40	37	74.0	263	1 GRAK_MOUSE	O35205 mus musculus
41	37	74.0	263	1 KLKR_PRANA	P32824 praomys nat
42	37	74.0	264	1 GRAK_HUMAN	P49863 homo sapien
43	37	74.0	277	1 KLKD_HUMAN	O9ukr3 homo sapien
44	36	72.0	73	1 EL2B_HORSE	P37358 equus cabal
45	36	72.0	226	1 DDN1_BOVIN	P80219 bos taurus

ALIGNMENTS

RESULT 1
KLK7_HUMAN
ID KLK7_HUMAN STANDARD; PRT; 253 AA.
AC P49862;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic
DE enzyme) (hSCCE).
GN KLK7 OR PRSS6 OR SCCE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.
RC TISSUE=Skin;
RX MEDLINE=94308225; PubMed=8034709;
RA Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A.,
RA Egelrud T.;
RT "Cloning, expression, and characterization of stratum corneum
RT chymotryptic enzyme. A skin-specific human serine proteinase.";
RL J. Biol. Chem. 269:19420-19426(1994).
[2]
RN SEQUENCE FROM N.A.
RP TISSUE=Keratinocytes;
RA Yousef G.M., Scoriilas A., Diamandis E.P.;
RT "Molecular characterization, mapping and tissue expression of the
RT human stratum corneum chymotryptic enzyme gene.";
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RX MEDLINE=20510030; PubMed=11054574;
RA Can L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
RA Moss P., Paepel B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene
RT cluster located in chromosome 19q13 region.";
RL Gene 257:119-130(2000).
[4]
RN CHARACTERIZATION.
RP MEDLINE=95314630; PubMed=7794273;
RX Skytt A., Stroemqvist M., Egelrud T.;
RA "Primary substrate specificity of recombinant human stratum corneum
RT chymotryptic enzyme.";
RL Biochem. Biophys. Res. Commun. 211:586-589(1995).
CC -!- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE
CC STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS
CC SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID
CC RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE
CC CLEAVES INSULIN B CHAIN AT 6-LEU-1-CYS-7, 16-TYR-1-LEU-17, 25-PHE-
CC 1-TYR-26, AND 26-TYR-1-THR-27. COULD PLAY A ROLE IN THE ACTIVATION
CC OF PRECURSORS TO INFLAMMATORY CYTOKINES.
CC -!- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS
CC EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE
CC ALSO SEEN IN THE BRAIN AND KIDNEY.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE

TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: L33404; AAC37551.1; -
EMBL: AF166330; AAD49718.1; -
EMBL: AF243527; AAG33360.1; -
HSSP: P00763; IDPO.
NEROPS: S01.300; -
MIM: 604438; -
InterPro: IPR001314; Chymotrypsin.
InterPro: IPR001254; Trypsin.
Pfam: PF00089; trypsin; 1.
PRINTS: PR00722; CHYMOTRYPSIN.
SMART: SM00020; TRYPSIN_DOM; 1.
PROSITE: PS02440; TRYPSIN_DOM; 1.
PROSITE: PS00134; TRYPSIN_HIS; 1.
PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; zymogen; glycoprotein; signal.
FT SIGNAL 1 22
FT PROPEP 23 29 ACTIVATION PEPTIDE.
FT CHAIN 30 253 KALLIKREIN 7.
FT ACT_SITE 70 70 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 112 112 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 205 205 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 36 137 BY SIMILARITY.
FT DISULFID 55 71 BY SIMILARITY.
FT DISULFID 137 239 BY SIMILARITY.
FT DISULFID 144 211 BY SIMILARITY.
FT DISULFID 176 190 BY SIMILARITY.
FT DISULFID 201 226 BY SIMILARITY.
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 253 AA; 27525 MW; 2D68B6B15A76A668 CRC64;

Query Match 100.0%; Score 50; DB 1; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.016;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 207 GPLVCRGTL 215
|||||

RESULT 2
KLK9_HUMAN STANDARD; PRT; 250 AA.
ID KLK9_HUMAN
AC Q9UKQ9;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kallikrein 9 precursor (EC 3.4.21.-) (Kallikrein-like protein 3) (KLK-L3).
GN KLK9.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20118156; PubMed=10652563;
RT Yousef G.M., Luo L.-Y., Diamandis E.P.;
RT "Identification of novel human kallikrein-like genes on chromosome 19q13.3-q13.4";
RL Anticancer Res. 19:2843-2852(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20247258; PubMed=10783266;

Yousef G.M., Diamandis E.P.;
"The expanded human kallikrein gene family: locus characterization and molecular cloning of a new member, KLK-L3";
Genomics 65:184-194(2000).
[3]
SEQUENCE FROM N.A.
MEDLINE=20510030; PubMed=11054574;
RX Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J., Moss P., Paepker B., Wang K.;
"Sequencing and expression analysis of the serine protease gene cluster located in chromosome 19q13 region.";
Gene 257:119-130(2000).
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- TISSUE SPECIFICITY: SKIN, THYMUS, TRACHEA, CEREBELLUM AND SPINAL CORD.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: AF135026; AAD26427.2; -
EMBL: AF243527; AAG33362.1; -
HSSP: P00763; IDPO.
MIM: 605504; -
InterPro: IPR001314; Chymotrypsin.
InterPro: IPR001254; Trypsin.
Pfam: PF00089; trypsin; 1.
PRINTS: PR00722; CHYMOTRYPSIN.
SMART: SM00020; TRYPSIN_DOM; 1.
PROSITE: PS02440; TRYPSIN_DOM; 1.
PROSITE: PS00134; TRYPSIN_HIS; 1.
PROSITE: PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; glycoprotein; signal.
FT SIGNAL 1 15 POTENTIAL.
FT CHAIN 16 250 KALLIKREIN 9.
FT ACT_SITE 63 63 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 111 111 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 204 204 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 29 164 BY SIMILARITY.
FT DISULFID 48 64 BY SIMILARITY.
FT DISULFID 136 238 BY SIMILARITY.
FT DISULFID 143 210 BY SIMILARITY.
FT DISULFID 175 189 BY SIMILARITY.
FT DISULFID 200 225 BY SIMILARITY.
FT CARBOHYD 131 131 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 250 AA; 27512 MW; F2785245B063E98B CRC64;

Query Match 90.0%; Score 45; DB 1; Length 250;
Best Local Similarity 88.9%; Pred. No. 0.14;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 206 GPLVCRGTL 214
|||||

RESULT 3
KLKE_HUMAN STANDARD; PRT; 251 AA.
ID KLKE_HUMAN
AC Q9P0G3;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kallikrein 14 precursor (EC 3.4.21.-) (Kallikrein-like protein 6) (KLK-L6).

GN KKL14 OR KKL16.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yousef G.M., Diamandis E.P.;
 RT "Molecular characterization, mapping, and tissue expression of KKL16,
 a hormonally regulated kallikrein-like gene.";
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
 Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
 Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Ganes J.,
 Dangnanan L., Eriker A., Christensen M., Georgescu A., Avila J., Liu S.,
 Andreise T., Trankhelm M., Attix C., Amico-Keller G., Coefield J.,
 Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
 Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
 Olsen A.S., Carrano A.V.;
 RT "Sequence analysis of chromosome 19q13.4.";
 RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=20545474; PubMed=10969073;
 RA Harvey T.J., Hooper J.D., Myers S.A., Stephenson S.A., Ashworth L.K.,
 Clements J.A.;
 RT "tissue-specific expression patterns and fine mapping of the human
 kallikrein (KLK) locus on proximal 19q13.4.";
 RL J. Biol. Chem. 275:37397-37406(2000).
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
 CC -1- TISSUE SPECIFICITY: HIGH EXPRESSION IN BRAIN, BONE MARROW AND
 FETAL LIVER. ALSO EXPRESSED IN LIVER, PANCREAS, FETAL SPLEN AND
 PROSTATE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF161221; AD50773.2; -;
 DR EMBL; AC011473; AAG23260.1; -;
 DR MEROPS; S01.029; -;
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease; Signal; Zymogen.
 FT SIGNAL 1 18
 FT PROPEP 19 24
 FT CHAIN 25 251
 FT ACT_SITE 67 67
 FT ACT_SITE 111 111
 FT ACT_SITE 204 204
 FT ACT_SITE 31 164
 FT DISULFID 52 68
 FT DISULFID 143 210
 FT DISULFID 175 189
 FT DISULFID 200 225
 FT SEQUENCE 251 AA; 27452 MW; 9087953BAFA7ED25 CRC64;
 Query Match 88.0%; Score 44; DB 1; Length 251;
 Best Local Similarity 88.9%; Pred. No. 0.22;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 GPLVCRCTL 9
 Db 206 GPLVCRGQL 214
 RESULT 4
 VSP3_BOTJA STANDARD; PRT; 258 AA.
 ID VSP3_BOTJA
 AC Q9PTU8;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Venom serine proteinase A precursor (EC 3.4.21.-).
 OS Bothrops jararaca (Jararaca).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
 OC Viperidae; Crotalinae; Bothrops.
 OX NCBI_TaxID=8724;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Venom gland.
 RA Murayama N.;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 TRYPSIN FAMILY. SNAKE VENOM SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB031394; BAA89310.1; -;
 DR HSP; P00763; IDPO.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 DR Hydrolase; Serine protease; Glycoprotein; Venom; Zymogen; Signal.
 FT SIGNAL 1 18
 FT PROPEP 19 24
 FT CHAIN 25 258
 FT ACT_SITE 65 65
 FT ACT_SITE 110 110
 FT ACT_SITE 204 204
 FT DISULFID 31 163
 FT DISULFID 50 66
 FT DISULFID 98 256
 FT DISULFID 142 210
 FT DISULFID 174 189
 FT DISULFID 200 225
 FT CARBOHYD 32 32
 FT CARBOHYD 44 44
 FT CARBOHYD 103 103
 FT CARBOHYD 121 121
 FT CARBOHYD 154 154
 FT CARBOHYD 170 170
 FT CARBOHYD 211 211
 FT CARBOHYD 251 251
 FT SEQUENCE 258 AA; 28058 MW; 034F49FDCB79EB64 CRC64;
 Query Match 80.0%; Score 40; DB 1; Length 258;

```
Best Local Similarity 75.0%; Pred. No. 1.3;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGT 8
    |||:|:|
Db 206 GPLICNGT 213

RESULT 5
NRPN_RAT NRPN_RAT STANDARD: PRT; 260 AA.
AC O88780;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neuropilin precursor (EC 3.4.21.-) (NP) (Kallikrein 8) (Brain serine
DE protease 1)..
GN KLK8 OR PRSS19 OR NRPN OR BSP1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=Brain;
RX MEDLINE=98389725; PubMed=9722524;
RA Davies B.J., Pickard B.S., Steel M., Morris R.G.M., Lathe R.;
RT "Serine proteases in rodent hippocampus.";
RL J. Biol. Chem. 273:23004-23011(1998).
CC -!- FUNCTION: SUGGESTED TO BE INVOLVED IN KINDLING EPILEPTOGENESIS AND
CC HIPPOCAMPAL PLASTICITY. HAS A STRONG PROTEOLYTIC ACTIVITY AGAINST
CC FIBRONECTIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -!- SUBCELLULAR LOCATION: Secreted (By similarity).
CC -!- TISSUE SPECIFICITY: RESTRICTED TO HIPPOCAMPUS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: AJ005641; CAA06643.1;
DR HSPG: Q61955; INPM.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS0240; TRYP_DOM; 1.
DR PROSITE: PS00134; TRYP_SIN_HIS; 1.
DR PROSITE: PS00135; TRYP_SIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 32 BY SIMILARITY.
FT CHAIN 33 260 NEUROPSIN.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 39 173 BY SIMILARITY.
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 145 246 BY SIMILARITY.
FT DISULFID 152 218 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 208 233 BY SIMILARITY.
FT CARBOHYD 110 110 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 260 AA; 28510 MW; 58DF4F0602A0B7F5 CRC64;

Query Match 80.0%; Score 40; DB 1; Length 260;

Best Local Similarity 77.8%; Pred. No. 1.3;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
    |||||:|
Db 214 GPLVCNGVL 222

RESULT 6
KLK2_HUMAN KLK2_HUMAN STANDARD: PRT; 261 AA.
AC P20151; Q15946; Q9U329;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glandular kallikrein 2 precursor (EC 3.4.21.35) (Tissue kallikrein)
DE (Prostate) (hGK-1).
GN KLK2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=88054467; PubMed=2824146;
RA Schedlich L.J., Bennetts B.H., Morris B.J.;
RT "Primary structure of a human glandular kallikrein gene.";
RL DNA 6:429-437(1987).
RN [2]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RP SEQUENCE=Prostate;
RX MEDLINE=92324494; PubMed=1726490;
RA Riegman P.H., Vlietstra R.J., der Korput H.A., Romijn J.C.,
RA Trapman J.;
RT "Identification and androgen-regulated expression of two major human
RT glandular kallikrein-1 (hGK-1) mRNA species.";
RL Mol. Cell. Endocrinol. 76:181-190(1991).
RN [3]
SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RX MEDLINE=20012757; PubMed=10544017;
RA Liu X.F., Essand M., Vasmatazis G., Lee B., Pastan I.;
RT "Identification of three new alternate human kallikrein 2 transcripts:
RT evidence of long transcript and alternative splicing.";
RL Biochem. Biophys. Res. Commun. 264:833-839(1999).
RN [4]
SEQUENCE FROM N.A.
RP SEQUENCE=Prostate;
RX MEDLINE=20510030; PubMed=11054574;
RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
RA Moss P., Paepker B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene
RT cluster located in chromosome 19q13 region.";
RL Gene 257:119-130(2000).
RN [5]
SEQUENCE FROM N.A.
RP SEQUENCE=Prostate;
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K.J., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Barnes J.,
RA Danganan L., Eriar A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Attix C., Andreise T., Frankheim M., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Kronmiller B.,
RA Arellano A., Brower A., Sanders C., Ow D., Nolan M., Trong S.,
RA Kobayashi A., Olsen A.S., Carrano A.V.;
RT "Sequence analysis of a 4.8 Mb region of 19q13.4 between KLK1 and
RT SYT3".
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of Arg-L-Xaa bonds in
CC small molecule substrates. Highly selective action to release
CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
CC Met-L-Xaa or Leu-L-Xaa.
CC -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2/PGK-10A AND 3;
CC ARE PRODUCED BY ALTERNATIVE SPLICING.
```

```

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY, KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M18156; ; NOT_ANNOTATED_CDS.
DR EMBL; M18157; AAA74454.1; -.
DR EMBL; S39329; AAD13816.1; -.
DR EMBL; S39329; AAD13817.1; -.
DR EMBL; AF188745; AAF08275.1; -.
DR EMBL; AF188746; AAF08276.1; -.
DR EMBL; AF188747; AAF08277.1; -.
DR EMBL; AF243527; AAG33356.1; -.
DR EMBL; AC037199; AAF63185.1; -.
DR PIR; A29586; A29586.
DR HSP; P07288; 1pfa.
DR MEROPS; S01.161; -.
DR MIM; 147960; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00240; TRYPsin_DOM; 1.
DR PROSITE; PS00134; TRYPsin_HIS; 1.
DR PROSITE; PS00135; TRYPsin_SER; 1.
DR KW Hydroxylase; Serine protease; Glycoprotein; Multigene family; Zymogen;
KW Signal; Alternative splicing. PROBABLE.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 261
FT ACT_SITE 65 65
FT ACT_SITE 120 120
FT ACT_SITE 213 213
FT DISULFID 31 173
FT DISULFID 50 66
FT DISULFID 152 219
FT DISULFID 184 198
FT DISULFID 209 234
FT CARBOHYD 102 102
FT VARSPIC 211 261
FT VARSPIC 165 261
FT SEQUENCE 261 AA; 28671 MW; 9CF7F4A1162EF42D CRC64;
SQ
Query Match 80.0%; Score 40; DB 1; Length 261;
Best Local Similarity 77.8%; Pred. No. 1.3;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
Db 215 GPLVCGVL 223
|||||
RESULT 7
KLK3_HUMAN
ID KLK3_HUMAN STANDARD; PRT; 261 AA.
AC P07288; Q16272;
DC 01-APR-1988 (Rel. 07, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Prostate specific antigen precursor (EC 3.4.21.77) (PSA) (Gamma-
DE semino protein) (kallikrein 3) (Semenogelase) (Seminin) (P-30 antigen).
GN KLK3 OR KLKB1 OR APS.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=89183632; PubMed=2467258;
RA Digby M.R., Zhang X.Y., Richard R.I.;
RT "Human prostate specific antigen (PSA) gene: structure and linkage to
RT the kallikrein-like gene, hCK-1."
RL Nucleic Acids Res. 17:2137-2137(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=89282407; PubMed=2471958;
RA Klobbeck H.-G., Combriato G., Schulz P., Arbusow V., Fittler F.;
RT "Genomic sequence of human prostate specific antigen (PSA).";
RL Nucleic Acids Res. 17:3981-3981(1989).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Leukocyte;
RX MEDLINE=89302090; PubMed=2472789;
RA Lundwall A.;
RT "Characterization of the gene for prostate-specific antigen, a human
RT glandular kallikrein."
RL Biochem. Biophys. Res. Commun. 161:1151-1159(1989).
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=87190978; PubMed=2436946;
RA Lundwall A., Lilja H.;
RT "Molecular cloning of human prostate specific antigen cDNA."
RL FEBS Lett. 214:317-322(1987).
RN [5]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=89246551; PubMed=2470373;
RA Henttu P., Viikko P.;
RT "cDNA coding for the entire human prostate specific antigen shows
RT high homologies to the human tissue kallikrein genes."
RL Biochem. Biophys. Res. Commun. 160:903-910(1989).
RN [6]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RX MEDLINE=89165891; PubMed=2466464;
RA Riegman P.H.J., Vlietstra R.J., van der Korput J.A.G.M., Romijn J.C.,
RA Trapman J.;
RT "Characterization of the prostate-specific antigen gene: a novel
RT human kallikrein-like gene."
RL Biochem. Biophys. Res. Commun. 159:95-102(1989).
RN [7]
RP SEQUENCE FROM N.A.
RC TISSUE=Prostate;
RA Monne M.M., Moreno J.M., Mele C.M., Mulholland G.M., Gomella L.G.;
RN Submitted (DEC-1994) to the EMBL/GenBank/DBJ databases.
RN [8]
RP SEQUENCE FROM N.A.
RX MEDLINE=20510030; PubMed=11054574;
RA Can L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
RA Moss P., Paepfer B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene
RT cluster located in chromosome 19q13 region."
RL Gene 257:119-130(2000).
RN [9]
RP SEQUENCE OF 1-176 FROM N.A.
RX MEDLINE=95079406; PubMed=7527295;
RA Monne M., Croce C.M., Yu H., Diamandis E.P.;
RT "Molecular characterization of prostate-specific antigen messenger RNA
RT expressed in breast tumors."
RL Cancer Res. 54:6344-6347(1994).
RN [10]
RP SEQUENCE OF 17-261 FROM N.A.
RX MEDLINE=88289366; PubMed=2456523;
RA Schulz P., Stucka R., Feldmann H., Combriato G., Klobbeck H.-G.,
RA Fittler F.;

```

RT "Sequence of a cDNA clone encompassing the complete mature human prostate specific antigen (PSA) and an unspliced leader sequence."; Nucleic Acids Res. 16:6226-6226(1988).

RN [11]

RX MEDLINE-86205857; PubMed-2422647;

RA Watt K.W.K., Lee P.J., M'Tinkulu T., Chan W.P., Loo R.;

RT "Human prostate-specific antigen: structural and functional similarity with serine proteases.";

RL Proc. Natl. Acad. Sci. U.S.A. 83:3166-3170(1986).

RN [12]

RP SEQUENCE OF 25-261.

RX MEDLINE-88082806; PubMed-3691515;

RA Schaller J., Akiyama K., Tsuda R., Hara M., Marti T., Rickli E.E.;

RT "Isolation, characterization and amino-acid sequence of gamma-seminoprotein, a glycoprotein from human seminal plasma.";

RL Eur. J. Biochem. 170:111-120(1987).

RN [13]

RP 3D-STRUCTURE MODELING.

RX MEDLINE-95218633; PubMed-7535613;

RA Villoutreix B.O., Gettoff E.D., Griffin J.H.;

RT "A structural model for the prostate disease marker, human prostate-specific antigen.";

RL Protein Sci. 3:2033-2044(1994).

RN [14]

RP 3D-STRUCTURE MODELING.

RX MEDLINE-98427950; PubMed-9751643;

RA Coombs G.S., Bergstrom R.C., Pellequer J.L., Baker S.I., Navre M., Smith M.M., Tainer J.A., Madison E.L., Corey D.R.;

RT "Substrate specificity of prostate-specific antigen (PSA)."; Chem. Biol. 5:475-488(1998).

RL Chem. Biol. 5:475-488(1998).

CC -1- FUNCTION: PRESUMABLY HYDROLYZE THE HIGH MOLECULAR MASS SEMINAL VESICLE PROTEIN THUS LEADING TO THE LIQUEFACTION OF THE SEMINAL COAGULUM.

CC -1- CATALYTIC ACTIVITY: Preferential cleavage: Tyr-I-Xaa.

CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: X14810; CAA32915.1; -

DR EMBL: X13940; CAA32123.1; -

DR EMBL: X13941; CAA32124.1; ALT_SEQ.

DR EMBL: X13942; CAB45487.1; -

DR EMBL: X13943; CAA32126.1; -

DR EMBL: X13944; CAA32127.1; -

DR EMBL: X05332; CAA28947.1; -

DR EMBL: X07730; -; NOT_ANNOTATED_CDS.

DR EMBL: M27274; AAA60192.1; -

DR EMBL: M26663; AAA58802.1; -

DR EMBL: U17040; AAA56764.1; -

DR EMBL: AF243527; AAG33355.1; -

DR EMBL: S75755; AAD14185.1; ALT_INIT.

DR EMBL: M24543; AAA60193.1; ALT_SEQ.

DR PIR: S03604; S03604.

DR PIR: A26757; A26757.

DR PIR: A32297; A32297.

DR PIR: A32423; A32423.

DR PDB: 1PFA; 26-JAN-95.

DR PDB: 2PSA; 16-MAR-99.

DR MEROPS: S01.162; -

DR MIM: 176820; -

DR InterPro: IPR001314; Chymotrypsin.

DR InterPro: IPR001254; Trypsin.

DR Pfam: PF00089; trypsin; 1

DR PRINTS: PR00722; CHYMOTRYPSIN.

DR SMART: SM00020; TRYP-SPC; 1.

DR PROSITE: PS0240; TRYPSIN_DOM; 1.

DR PROSITE: PS00134; TRYPSIN_HIS; 1.

DR PROSITE: PS00135; TRYPSIN_SER; 1.

KW Hydrolase; Serine protease; Antigen; Zymogen; Signal;

KW 3D-structure.

FT SIGNAL 1 18 POTENTIAL.

FT PROPEP 19 24 ACTIVATION PEPTIDE.

FT CHAIN 25 261 PROSTATE SPECIFIC ANTIGEN.

FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).

FT CARBOHYD 69 69 N-LINKED (GLCNAC...).

FT DISULFID 31 173 BY SIMILARITY.

FT DISULFID 50 66 BY SIMILARITY.

FT DISULFID 152 219 BY SIMILARITY.

FT DISULFID 184 198 BY SIMILARITY.

FT DISULFID 209 234 BY SIMILARITY.

FT CONFLICT 64 64 A -> T (IN REF. 10).

FT CONFLICT 73 73 I -> T (IN REF. 5).

FT CONFLICT 86 86 T -> I (IN REF. 5).

FT CONFLICT 94 94 H -> T (IN REF. 11).

FT CONFLICT 136 136 V -> M (IN REF. 10).

FT CONFLICT 165 168 FLTP -> HLLYDQM (IN REF. 11).

FT CONFLICT 175 175 D -> Q (IN REF. 11).

FT CONFLICT 175 175 D -> P (IN REF. 5).

FT CONFLICT 184 184 C -> Q (IN REF. 5).

FT CONFLICT 260 260 N -> D (IN REF. 5).

SQ SEQUENCE 261 AA; AE9E732AF872141A CRC64;

Query Match 80.0%; Score 40; DB 1; Length 261;

Best Local Similarity 77.8%; Pred. No. 1.3;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

DB 215 GPLVCGNVL 223

RESULT 8

KLK1_RAT STANDARD; PRT; 261 AA.

ID KLK1_RAT

AC P00758;

DT 21-JUL-1986 (Rel. 01, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 15-DEC-1998 (Rel. 37, Last annotation update)

DE Glandular kallikrein, pancreatic 1 precursor (EC 3.4.21.35) (Tissue

DE kallikrein) (pS kallikrein) (MGK-1).

GN KLK1 OR KLK-1.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE-83117659; PubMed=6961406;

RA Swift G.H., Dagorn J.-C., Ashley P.L., Cummings S.W., McDonald R.J.;

RT "Rat pancreatic kallikrein mRNA: nucleotide sequence and amino acid

RL sequence of the encoded preproenzyme.";

RL Proc. Natl. Acad. Sci. U.S.A. 79:7263-7267(1982).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE-86051477; PubMed=2998455;

RA Ashley P.L., Macdonald R.J.;

RT "kallikrein-related mRNAs of the rat submaxillary gland: nucleotide

RL sequences of four distinct types including tonin.";

RL Biochemistry 24:4512-4520(1985).

RN [3]

RP SEQUENCE FROM N.A.

RC TISSUE-Kidney;

RX MEDLINE-89327211; PubMed=2753879;

RA Inoue H., Fukui K., Miyake Y.;

RT "Identification and structure of the rat true tissue kallikrein gene

expressed in the kidney.";
J. Biochem. 105:834-840(1989).
[4]
SEQUENCE FROM N.A.
MEDLINE=89214217; PubMed=2708383;
Wines D.R., Brady J.M., Pritchett D.B., Roberts J.L., Macdonald R.J.;
"Organization and expression of the rat kallikrein gene family.";
J. Biol. Chem. 264:7653-7662(1989).
[5]
SEQUENCE OF 48-261 FROM N.A.
MEDLINE=86131678; PubMed=3004582;
Gerald W.L., Chao J., Chao L.;
"Immunological identification of rat tissue kallikrein cDNA and
characterization of the kallikrein gene family.";
Biochim. Biophys. Acta 866:1-14(1986).
CC -|- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -|- CATALYTIC ACTIVITY: Preferential cleavage of Arg-|-Xaa bonds in
small molecule substrates. Highly selective action to release
kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
Met-|-Xaa or Leu-|-Xaa.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; J00758; -. NOT_ANNOTATED.CDS.
DR EMBL; M11563; AAA41464.1; ALT_INIT.
DR EMBL; M23876; AAA41462.1; -.
DR EMBL; M23874; AAA41462.1; JOINED.
DR EMBL; M23875; AAA41462.1; JOINED.
DR EMBL; D00446; BAA00346.1; ALT_INIT.
DR EMBL; D00448; BAA00346.1; JOINED.
DR EMBL; D00447; BAA00346.1; JOINED.
DR EMBL; X03560; CAA27247.1; -.
DR PIR; A00944; KQTP.
DR PIR; A23863; A23863.
DR PIR; JX0073; JX0073.
DR PIR; A33359; A33359.
DR HSP; P00757; ISGF.
DR MEROPS; S01.405; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR K0 Hydrolyase; Serine protease; Glycoprotein; Multigene family; Zymogen;
KW Signal.
FT SIGNAL 1 18 PROBABLE.
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
FT CHAIN 25 261 GLANDULAR KALLIKREIN 1.
FT CHAIN 25 111 CHAIN 1.
FT CHAIN 112 261 CHAIN 2.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT DISULFID 31 173 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 152 219 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 209 234 BY SIMILARITY.
FT CARBOHYD 108 108 N-LINKED (GLCNAC...) (PROBABLE).
FT SEQUENCE 261 AA; 28852 MW; F2F99C0227A7882B CRC64;

Query Match 78.0%; Score 39; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 2.1;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRCTL 9
Db 215 GPLTCNGVL 223
| | | : | | |

RESULT 9
KLKC_HUMAN STANDARD; PRT; 248 AA.
ID KLKC_HUMAN Q9UKR0; Q9UKR1;
AC Q9UKR0; Q9UKR1;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kallikrein 12 precursor (EC 3.4.21.-) (Kallikrein-like protein 5)
DE (KLK-L5).
GN KLK12 OR KLK15.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
[1]
RN RP SEQUENCE FROM N.A. (ISOFORM 1).
RA MEDLINE=20118156; PubMed=10652563;
RX Yousef G.M., Luo L.-Y., Diamandis E.P.;
RT "Identification of novel human kallikrein-like genes on chromosome
RT 19q13.3-q13.4.";
RL Anticancer Res. 19:2843-2852(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RA Yousef G.M., Magklara A., Scorilas A., Diamandis E.P.;
RT "Cloning of new alternatively spliced forms of the kallikrein-like
RT gene 5 (KLK-L5).";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=20510030; PubMed=11054574;
RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
RA Moss P., Paepker B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene
RT cluster located in chromosome 19q13 region.";
RL Gene 257:119-130(2000).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RA Lamerdin J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Garnes J.,
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Andreise T., Trankheim M., Attix C., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of chromosome 19q13.4.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: Secreted (Probable).
CC produced by alternative splicing.
CC -|- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2; are
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; AF135025; AAD26426.2; -.
DR EMBL; AF135025; AAF06065.1; -.
DR EMBL; AF243527; AAG33365.1; -.
DR

```
DR EMBL; AC011473; AAC23258.1; -.
DR HSPP; P00761; IAKS.
DR MEROPS; S01.020; -.
DR MIM; 605539; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Signal;
KW Alternative splicing.
FT SIGNAL 1 17 POTENTIAL.
FT CHAIN 18 248 KALLIKREIN 12.
FT ACT_SITE 62 62 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 108 108 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 200 200 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 28 161 BY SIMILARITY.
FT DISULFID 47 63 BY SIMILARITY.
FT DISULFID 133 235 BY SIMILARITY.
FT DISULFID 140 206 BY SIMILARITY.
FT DISULFID 172 186 BY SIMILARITY.
FT DISULFID 196 222 BY SIMILARITY.
FT CARBOHYD 24 24 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 163 163 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPPLIC 236 248 KYVDWTRIMRN -> NSTLVGLGTSWNFNSQPF (IN ISOFORM 2).
SQ SEQUENCE 248 AA; 26733 MW; BB473E98F8BAF703 CRC64;

Query Match 76.0%; Score 38; DB 1; Length 248;
Best Local Similarity 77.8%; Pred. No. 3.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
    |||||
Db 202 GPLVCGGVL 210

RESULT 10
CFAD_HUMAN STANDARD; PRT; 253 AA.
AC P00746;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Complement factor D precursor (Ec 3.4.21.46) (C3 convertase activator)
DE (Properdin factor D) (Adipsin).
DE DF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE-92250520; PubMed-1374388;
RA White R.T., Damm D., Hancock N., Rosen B.S., Lowell B.B., Usher P.,
RA Flier J.S., Spiegelman B.M.;
RT "Human adipsin is identical to complement factor D and is expressed
at high levels in adipose tissue.";
RL J. Biol. Chem. 267:9210-9213(1992).
RN [2]
RP SEQUENCE FROM N.A.
RA Flier J.S., Spiegelman B.M., Rosen B.M.;
RL Patent number WO9006365, 14-JUN-1990.
RN [3]
RP SEQUENCE OF 26-252.
RX MEDLINE-85000441; PubMed-6383466;
RA Niemann M.A., Bhown A.S., Bennett J.C., Volanakis J.E.;
RT "Amino acid sequence of human D of the alternative complement
pathway.";
RL Biochemistry 23:2482-2486(1984).
```

```
RN [4]
RP PARTIAL SEQUENCE OF 26-252.
RX MEDLINE-84108950; PubMed-6363133;
RA Johnson D.M.A., Gagnon J., Reid K.B.M.;
RT "Amino acid sequence of human factor D of the complement system.
Similarity in sequence between factor D and proteases of non-plasma
origin.";
RL FEBS Lett. 166:347-351(1984).
RN [5]
RP PARTIAL SEQUENCE OF 26-61 AND 194-220.
RX MEDLINE-84256515; PubMed-6821372;
RA Johnson D.M.A., Gagnon J., Reid K.B.M.;
RT "Factor D of the alternative pathway of human complement.
Purification, alignment and N-terminal amino acid sequences of the
major cyanogen bromide fragments, and localization of the serine
residue at the active site.";
RL Biochem. J. 187:863-874(1980).
RN [6]
RP PARTIAL SEQUENCE OF 26-82.
RX MEDLINE-80145719; PubMed-6987665;
RA Volanakis J.E., Bhown A.S., Bennett J.C., Mole J.E.;
RT "Partial amino acid sequence of human factor D:homology with serine
proteases.";
RL Proc. Natl. Acad. Sci. U.S.A. 77:1116-1119(1980).
RN [7]
RP PARTIAL SEQUENCE OF 26-78.
RX MEDLINE-81054886; PubMed-6776531;
RA Davis A.E. III;
RT "Active site amino acid sequence of human factor D.";
RL Proc. Natl. Acad. Sci. U.S.A. 77:4938-4942(1980).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE-94118317; PubMed-828289;
RA Narayana S.V.L., Carson M., El-Kabbani O., Kilpatrick J.M., Moore D.,
RA Chen X., Bugg C.E., Volanakis J.E., Delucas L.J.;
RT "Structure of human factor D. A complement system protein at 2.0-A
resolution.";
RL J. Mol. Biol. 235:695-708(1994).
RN [9]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE-96025834; PubMed-7592653;
RA Kim S., Narayana S.V., Volanakis J.E.;
RT "Crystal structure of a complement factor D mutant expressing
enhanced catalytic activity.";
RL J. Biol. Chem. 270:24399-24405(1995).
CC !- FUNCTION: FACTOR D CLEAVES FACTOR B WHEN THE LATTER IS COMPLEXED
CC WITH FACTOR C3B, ACTIVATING THE C3BBB COMPLEX, WHICH THEN BECOMES
CC THE C3 CONVERTASE OF THE ALTERNATE PATHWAY. ITS FUNCTION IS
CC HOMOLOGOUS TO THAT OF C1S IN THE CLASSICAL PATHWAY.
CC !- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
CC !- CAUTION: IN ADDITION TO THE CONFLICTS SHOWN IN THE FEATURE TABLE,
CC REF.3 SEQUENCE HAD A PEPTIDE IN THE WRONG ORDER AND ANOTHER ONE
CC MISSING.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M84526; AAA35527.1; ALT_INIT.
DR PIR; A40197; DBHU.
DR PDB; IDFP; 25-FEB-98.
DR PDB; IDST; 11-JUL-96.
DR PDB; IDSU; 17-AUG-96.
DR MEROPS; S01.191; -.
DR MIM; 134350; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
```

```
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_Spc; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Complement alternate pathway; Plasma; Hydrolase; Serine protease;
FT Zymogen; Glycoprotein; Signal; 3D-structure.
FT SIGNAL 1 20
FT PROPEP 21 25
FT CHAIN 26 25
FT ACT_SITE 66
FT ACT_SITE 112 112
FT ACT_SITE 208 208
FT DISULFID 51 67
FT DISULFID 148 214
FT DISULFID 179 195
FT DISULFID 204 229
FT CONFLICT 26 26
FT CONFLICT 35 35
FT CONFLICT 40 40
FT CONFLICT 49 49
FT CONFLICT 52 52
FT CONFLICT 59 59
FT CONFLICT 63 63
FT CONFLICT 73 73
FT CONFLICT 83 86
FT CONFLICT 83 84
FT CONFLICT 94 95
FT CONFLICT 96 96
FT CONFLICT 136 136
FT CONFLICT 178 191
FT CONFLICT 243 243
FT CONFLICT 250 250
FT CONFLICT 250 250
SQ SEQUENCE 253 AA; 27004 MW; BD553B70BD55C6AD CRC64;

Query Match 76.0%; Score 38; DB 1; Length 253;
Best Local Similarity 77.8%; Pred. No. 3.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
| | | | |
DB 210 GPLVCGGVL 218

RESULT 11
VSP3_TRIGA STANDARD; PRT; 258 AA.
AC O13063;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Venom serine proteinase 3 precursor (EC 3.4.21.-).
GN TLG3.
OS Trimeresurus gramineus (Indian green tree viper) (Green habu snake).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
OC Viperidae; Crotalinae; Trimeresurus.
OX NCBI_TaxID=8767;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Venom gland;
RX MEDLINE=97096898; PubMed=8941719;
RA Deshmara M., Ogawa T., Nakashima K., Nobuhisa I., Chijiwa T.,
RA Shimohigashi Y., Fukumaki Y., Niwa M., Yamashina I., Hattori S.,
RA Ohno M.;
RT "Accelerated evolution of crotalinae snake venom gland serine
proteases.";
RL FEBS Lett. 397:83-88(1996).
CC -1- FUNCTION: THROMBIN-LIKE SNAKE VENOM SERINE PROTEASE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: EXPRESSED BY VENOM GLAND.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
```

```
CC TRYPsin FAMILY. SNAKE VENOM SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; D67085; BAA19983.1; -.
CC HSP; P00763; IDPO.
CC MEROPS; S01.185; -.
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Trypsin.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; Tryp_Spc; 1.
CC PROSITE; PS00240; TRYPSIN_DOM; 1.
CC PROSITE; PS00134; TRYPSIN_HIS; 1.
CC PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Venom; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 258
FT ACT_SITE 65 65
FT ACT_SITE 110 110
FT ACT_SITE 204 204
FT DISULFID 31 163
FT DISULFID 98 256
FT DISULFID 142 210
FT DISULFID 174 189
FT DISULFID 200 225
FT CARBOHYD 44 44
FT CARBOHYD 103 103
FT CARBOHYD 117 117
FT CARBOHYD 121 121
FT CARBOHYD 154 154
FT CARBOHYD 251 251
SQ SEQUENCE 258 AA; 28034 MW; 3160379F61E9F74B CRC64;

Query Match 76.0%; Score 38; DB 1; Length 258;
Best Local Similarity 66.7%; Pred. No. 3.2;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
| | | | |
DB 206 GPLICNGQL 214

RESULT 12
KLK8_HUMAN STANDARD; PRT; 260 AA.
AC O60259; Q9U047; Q9HCB3; Q9U1L9;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neuropsin precursor (EC 3.4.21.-) (NP) (Kallikrein 8) (Ovasin) (Serine
DE protease TADG-14) (Tumor-associated differentially expressed gene-14
DE protein).
GN KLK8 OR PRSSI9 OR TADG14 OR NRPN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Hippocampus;
RX MEDLINE=98372070; PubMed=9714609;
RA Yoshida S., Taniguchi M., Hirata A., Shiosaka S.;
RT "Sequence analysis and expression of human neuropsin cDNA and gene-14
RL Gene 213:9-16(1998).
```

RA [2] SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RP TISSUE=Brain;
RX MEDLINE=99203457; PubMed=10102990;
RA Mitsui S., Tsuruoka N., Yamashiro K., Nakazato H., Yamaguchi N.;
RT "A novel form of human neurospisin, a brain-related serine protease, is
RT generated by alternative splicing and is expressed preferentially in
RT human adult brain.";
RL Eur. J. Biochem. 260:627-634(1999).
RN [3]
RA [3] SEQUENCE FROM N.A. (ISOFORM 1).
RP TISSUE=Ovary;
RX MEDLINE=99413504; PubMed=10485494;
RA Underwood L.J., Tanimoto H., Wang Y., Shigemasa K., Parmley T.H.,
RA O'Brien T.J.;
RT "Cloning of tumor-associated differentially expressed gene-14, a novel
RT serine protease overexpressed by ovarian carcinoma.";
RL Cancer Res. 59:4439(1999).
RN [4]
RA [4] SEQUENCE FROM N.A. (ISOFORM 1).
RP Gan L., Gellinas R., Gown A.M., Moss P., Smith R., Wang K.;
RA "Molecular cloning and characterization of a novel serine protease,
RT ovasin, a potential molecular marker for ovarian carcinomas.";
RL Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
RN [5]
RA [5] SEQUENCE FROM N.A. (ISOFORM 1).
RP MEDLINE=20510030; PubMed=11054574;
RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
RA Moss P., Paepel B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene
RT cluster located in chromosome 19q13 region.";
RL Gene 257:119-130(2000).
RN [6]
RA [6] SEQUENCE OF 1-164 FROM N.A. (ISOFORM 1).
RP Lanerding J.E., McCready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stillwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Garnes J.,
RA Dangnanan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Andreise T., Trankheim M., Attix C., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carraro A.V.;
RT "Sequence analysis of chromosome 19q13.4.";
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RA [6] FUNCTION: SUGGESTED TO BE INVOLVED IN KINDLING EPILEPTOGENESIS AND
CC HIPPOCAMPAL PLASTICITY.
CC [6] CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC [6] SUBCELLULAR LOCATION: Secreted.
CC [6] ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC [6] TISSUE SPECIFICITY: ISOFORM 1 IS PREDOMINANTLY EXPRESSED IN THE
CC PANCREAS WHILE ISOFORM 2 IS EXPRESSED IN ADULT BRAIN AND
CC HIPPOCAMPUS. BOTH FORMS ARE ALSO FOUND IN FETAL BRAIN AND
CC PLACENTA. NOT DETECTED IN KIDNEY, SPLEEN, LIVER AND LUNG.
CC [6] SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB009849; BAA28673.1; -
CC EMBL; AB012761; BAA28676.1; -
CC EMBL; AB010780; BAA8684.1; -
CC EMBL; AB008390; BAA82665.1; -
CC EMBL; AB008927; BAA82666.1; -
CC EMBL; AF055982; AAD56050.1; -
CC EMBL; AF095742; AAD25979.1; -
CC EMBL; AF095743; AAD29574.1; -

DR EMBL; AF243527; AAG33361.1; -
DR EMBL; AC011473; AAG23254.1; -
DR HSSP; P00760; IAQ7.
DR MEROPS; S01.244; -.
DR MIM; 605644; -.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW Alternative splicing.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 32 BY SIMILARITY.
FT CHAIN 33 260 NEUROPSIN.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 39 173 BY SIMILARITY.
FT DISULFID 58 74 BY SIMILARITY.
FT DISULFID 145 246 BY SIMILARITY.
FT DISULFID 152 218 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 208 233 BY SIMILARITY.
FT CARBOHYD 110 110 N-LINKED (GLCNAC...) (POTENTIAL).
FT VARSPPLIC 23 23 WRSNPLPAA (IN ISOFORM 2).
FT SQ SEQUENCE 260 AA; 28048 MW; EF4395B8C83E660 CRC64;
Query Match 76.08; Score 38; DB 1; Length 260;
Best Local Similarity 77.88; Pred. No. 3.2;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCRGRTL 9
| | | | | | | | | |
DB 214 GPLVCDGAL 222
RESULT 13
KLK3_MACMU STANDARD; PRT; 261 AA.
ID KLK3_MACMU
AC P33619;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Prostate specific antigen precursor (EC 3.4.21.35) (PSA) (Gamma-
DE seminoprotein) (Kallikrein 3).
GN KLK3 OR APS
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;
OC Cercopitheidae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93363642; PubMed=7689340;
RA Gauthier E.R., Chapdelaine P., Tremblay R.R., Dube J.Y.;
RT "Characterization of rhesus monkey prostate specific antigen cDNA.";
RL Biochim. Biophys. Acta 1174:207-210(1993).
CC [1] FUNCTION: PSA PRESUMABLY HYDROLYZE THE HIGH MOLECULAR MASS
CC SEMINAL VESICLE PROTEIN THUS LEADING TO THE LIQUEFACTION OF
CC THE SEMINAL COAGULUM. PSA IS A GLANDULAR KALLIKREIN.
CC [1] CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: TYR-I-XAA.
CC [1] SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; X73560; CAA51957.1; -;
CC PIR; S34239; S34239.
CC PIR; S35711; S35711.
CC HSSP; P07288; 1PFA.
CC MEROPS; S01.162; -;
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Glycoprotein; Antigen; Zymogen; Signal.
FT SIGNAL 1 18
FT PROPEP 19 24
FT CHAIN 25 261
FT ACT_SITE 65 65
FT ACT_SITE 120 120
FT ACT_SITE 213 213
FT DISULFID 31 173
FT DISULFID 50 66
FT DISULFID 152 219
FT DISULFID 184 198
FT DISULFID 209 234
FT CARBOHYD 102 102
SQ SEQUENCE 261 AA; 28816 MW; 8525814B15967E5C CRC64;

Query Match 76.0%; Score 38; DB 1; Length 261;
Best Local Similarity 77.8%; Pred. No. 3.2;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 GPLVCRGTL 9
Db 215 GPLVCDGVL 223
| | | | | | | | | |

RESULT 14
KLK5_HUMAN STANDARD; PRT; 293 AA.
ID Q9V337; Q9HBC8;
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Kallikrein 5 precursor (EC 3.4.21.-) (Stratum corneum tryptic enzyme)
DE (Kallikrein-like protein 2) (KLK-L2).
GN KLK5 OR SCTE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Stratum corneum;
RX MEDLINE=99445563; PubMed=10514489;
RA Brattand M., Egelrud T.;
RT "Purification, molecular cloning, and expression of a human stratum
RT corneum trypsin-like serine protease with possible function in
RT desquamation.";
RL J. Biol. Chem. 274:30033-30040(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20118156; PubMed=10652563;
RA Yousef G.M., Luo L.-Y., Diamandis E.P.;
RT "Identification of novel human kallikrein-like genes on chromosome
RT 19q13.3-q13.4";
RL Anticancer Res. 19:2843-2852(1999).
RN [3]
RP SEQUENCE FROM N.A.

RX MEDLINE=20510030; PubMed=11054574;
RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
RA Moss P., Paepel B., Wang K.;
RT "Sequencing and expression analysis of the serine protease gene
RT cluster located in chromosome 19q13 region.";
RL Gene 257:119-130(2000).
CC -!- FUNCTION: MAY BE INVOLVED IN DESQUAMATION.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- TISSUE SPECIFICITY: EXPRESSED IN SKIN, BREAST, BRAIN AND TESTIS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL; AF168768; AAP03101.1; -;
CC EMBL; AF135028; AAD26429.1; -;
CC EMBL; AF243527; AAG33358.1; -;
CC HSSP; P00763; 1DPO.
CC MEROPS; S01.017; -;
CC MIN; G05643; -;
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolyase; Serine protease; Glycoprotein; Signal.
FT SIGNAL 1 22
FT CHAIN 23 293
FT ACT_SITE 108 108
FT ACT_SITE 153 153
FT ACT_SITE 245 245
FT DISULFID 73 206
FT DISULFID 93 109
FT DISULFID 178 279
FT DISULFID 185 251
FT DISULFID 217 231
FT DISULFID 241 266
FT CARBOHYD 69 69
FT CARBOHYD 173 173
FT CARBOHYD 208 208
FT CARBOHYD 252 252
FT CONFLICT 25 56
SQ SEQUENCE 293 AA; 32020 MW; D92C92F5609E5946 CRC64;

Query Match 76.0%; Score 38; DB 1; Length 293;
Best Local Similarity 66.7%; Pred. No. 3.6;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Oy 1 GPLVCRGTL 9
Db 247 GPVVCNGSL 255
| | | | | | | | | |

RESULT 15
CTA3_HUMAN STANDARD; PRT; 1154 AA.
ID CTA3_HUMAN
AC Q9BZ76; Q9C0E9;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Contactin associated protein-like 3 precursor (cell recognition
DE molecule Caspr3).
GN CNTNAP3 OR CASPR3 OR KIAA1714.
OS Homo sapiens (Human).


```
FT SIGNAL 1 25
FT CHAIN 26 1308
FT DOMAIN 26 1241
FT TRANSMEM 1242 1262
FT DOMAIN 1263 1308
FT DOMAIN 34 174
FT DOMAIN 212 364
FT DOMAIN 398 547
FT DOMAIN 549 586
FT DOMAIN 821 957
FT DOMAIN 958 997
FT DOMAIN 1046 1202
FT CARBOHYD 260 260
FT CARBOHYD 285 285
FT CARBOHYD 359 359
FT CARBOHYD 538 538
FT CARBOHYD 574 574
FT CARBOHYD 602 602
FT CARBOHYD 625 625
FT CARBOHYD 637 637
FT CARBOHYD 706 706
FT CARBOHYD 748 748
FT CARBOHYD 1023 1023
FT CARBOHYD 1073 1073
SQ SEQUENCE 1308 AA; 145316 MW; 145316 MW; DA2FF6C2C8B082B0 CRC64;

Query Match 76.0%; Score 38; DB 1; Length 1308;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
   |||:||||
Db 781 GPLLCRG 787

RESULT 17
CTA4_MOUSE
ID CTA4_MOUSE STANDARD; PRT; 1310 AA.
AC Q99P47;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Contactin associated protein-like 4 precursor (Cell recognition molecule Caspr4).
GN CYNAP4 OR CASPR4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Spiegel I., Schaeren-Wiemers N., Peles E.;
RT "Identification of two new members of the Caspr family.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
CC -|- SUBCELLULAR LOCATION: Type I membrane protein (Potential).
CC -|- SIMILARITY: CONTAINS 1 F5/8 TYPE C DOMAIN.
CC -|- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -|- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -|- SIMILARITY: BELONGS TO THE NEUREXIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF333770; AAC52890.1; -
CC InterPro; IPR000561; EGF-like.
CC InterPro; IPR000421; FA58.C.
CC InterPro; IPR002181; Fibrinogen_C.
```

```
DR InterPro; IPR001791; Laminin_G.
DR Pfam; PF00008; EGF_2.
DR Pfam; PF00754; F5_F8_type_C; 1.
DR Pfam; PF00054; laminin_G; 3.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00001; EGF_like; 2.
DR SMART; SM00231; FA58C; 1.
DR SMART; SM00186; FBG; 1.
DR SMART; SM00282; LamG; 4.
DR PROSITE; PS00022; EGF_1; FALSE_NEG.
DR PROSITE; PS01186; EGF_2; FALSE_NEG.
DR PROSITE; PS01285; FA58C_1; 1.
DR PROSITE; PS01286; FA58C_2; 1.
DR PROSITE; PS50025; LAM_G_DOMAIN; 4.
KW Glycoprotein; Cell adhesion; Signal; Transmembrane; Repeat.
FT SIGNAL 1 27
FT CHAIN 28 1310
FT DOMAIN 28 1243
FT TRANSMEM 1244 1264
FT DOMAIN 1265 1310
FT DOMAIN 37 176
FT DOMAIN 214 346
FT DOMAIN 400 529
FT DOMAIN 551 588
FT DOMAIN 823 960
FT DOMAIN 960 999
FT DOMAIN 1048 1204
FT CARBOHYD 262 262
FT CARBOHYD 287 287
FT CARBOHYD 361 361
FT CARBOHYD 540 540
FT CARBOHYD 576 576
FT CARBOHYD 604 604
FT CARBOHYD 627 627
FT CARBOHYD 639 639
FT CARBOHYD 708 708
FT CARBOHYD 750 750
FT CARBOHYD 1019 1019
FT CARBOHYD 1025 1025
FT CARBOHYD 1075 1075
SQ SEQUENCE 1310 AA; 144713 MW; BA34BEE91BE0FDB7 CRC64;

Query Match 76.0%; Score 38; DB 1; Length 1310;
Best Local Similarity 85.7%; Pred. No. 15;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
   |||:||||
Db 783 GPLLCRG 789

RESULT 18
KLK3_RAT
ID KLK3_RAT STANDARD; PRT; 188 AA.
AC P15950;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Glandular kallikrein 3, submandibular (EC 3.4.21.35) (Tissue kallikrein) (S1 kallikrein) (RGK-3) (RSGK-50) (Fragment).
GN KLK3 OR KLK-3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Shai S.Y., Woodley-Miller C., Chao J., Chao L.;
RT "Characterization of genes encoding rat tonin and a kallikrein-like serine protease.";
RL Biochemistry 28:5334-5343(1989).
RN [2]
```

RP SEQUENCE OF 33-188 FROM N.A.
 RA MEDLINE=86051477; PubMed=2998455;
 RX Ashley P.L., Macdonald R.J.;
 RT "Kallikrein-related mRNAs of the rat submaxillary gland: nucleotide
 sequences of four distinct types including tonin.";
 RL Biochemistry 24:4512-4520(1985).
 CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
 CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-Xaa bonds in
 CC small molecule substrates. Highly selective action to release
 CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
 CC Met-I-Xaa or Leu-I-Xaa.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M25534; AAA2080.1; -;
 DR EMBL; M11564; AAA41465.1; -;
 DR PIR; B23863; B23863.
 DR PIR; B23340; B23340.
 DR HSSP; P00759; ITON.
 DR MEROPS; S01.160; -;
 DR InterPro: IPR001254; Trypsin.
 DR Pfam; PF00089; Trypsin; 1.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS02040; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; PARTIAL.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Multigene family.
 FT NON_TER 1 1
 FT ACT_SITE 47 47 CHARGE RELAY SYSTEM.
 FT ACT_SITE 140 140 CHARGE RELAY SYSTEM.
 FT DISULFID 79 146 BY SIMILARITY.
 FT DISULFID 111 125 BY SIMILARITY.
 FT DISULFID 136 161 BY SIMILARITY.
 FT CONFLICT 34 35 MR -> IW (IN REF. 2).
 FT CONFLICT 42 42 K -> G (IN REF. 2).
 FT CONFLICT 186 186 E -> K (IN REF. 1).
 SQ SEQUENCE 188 AA; 20986 MW; B3ECEC65582EB1E9 CRC64;
 Query Match 74.0%; Score 37; DB 1; Length 188;
 Best Local Similarity 66.7%; Pred. No. 3.7;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GPLVCRGTL 9
 DB 142 GPLICDGL 150
 RESULT 19
 CUTL_ALTBK STANDARD; PRT; 209 AA.
 ID CUTL_ALTBK STANDARD; PRT; 209 AA.
 AC P41744;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Cytinase precursor (EC 3.1.1.74).
 GN CUTAB1.
 OS Alternaria brassicicola.
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Dothideomycetes;
 OC Pleosporales; Pleosporaceae; mitosporic Pleosporaceae; Alternaria.
 OX NCBI_TaxID=29001;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Yao C., Koeller W.;
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: CATALYZE THE HYDROLYSIS OF CUTIN, A POLYESTER THAT FORMS
 CC THE STRUCTURE OF PLANT CUTICLE. ALLOWS PATHOGENIC FUNGI TO
 CC PENETRATE THROUGH THE CUTICULAR BARRIER INTO THE HOST PLANT DURING
 CC THE INITIAL STAGE OF THE FUNGAL INFECTION (By similarity).
 CC -1- CATALYTIC ACTIVITY: Cutin + H(2)O = cutin monomers.
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC -1- SIMILARITY: BELONGS TO THE CUTINASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U03393; AAA03470.1; -;
 DR HSSP; P00590; 1XZG.
 DR InterPro: IPR000675; Cutinase.
 DR Pfam; PF01083; Cutinase; 1.
 DR PRINTS; PR00129; CUTINASE.
 DR PROSITE; PS00155; CUTINASE_1; 1.
 DR PROSITE; PS00931; CUTINASE_2; 1.
 KW Hydrolase; Serine esterase; Signal.
 FT SIGNAL 1 ? POTENTIAL.
 FT CHAIN ? 209 CUTINASE.
 FT DISULFID 29 177 BY SIMILARITY.
 FT DISULFID 108 170 BY SIMILARITY.
 FT ACT_SITE 119 119 BY SIMILARITY.
 FT ACT_SITE 174 174 BY SIMILARITY.
 FT ACT_SITE 187 187 BY SIMILARITY.
 SQ SEQUENCE 209 AA; 21648 MW; 1756D06D84093A64 CRC64;
 Query Match 74.0%; Score 37; DB 1; Length 209;
 Best Local Similarity 77.8%; Pred. No. 4.1;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GPLVCRGTL 9
 DB 173 GDLVCNGTL 181
 RESULT 20
 KLK2_CAVPO
 ID KLK2_CAVPO STANDARD; PRT; 239 AA.
 AC P12323;
 DT 01-OCT-1989 (Rel. 12, Created)
 DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glandular kallikrein, prostatic (EC 3.4.21.35) (Tissue kallikrein)
 DE (Prostate esterase).
 OS Cavia porcellus (Guinea pig).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.
 OX NCBI_TaxID=10141;
 RN [1]
 RP SEQUENCE
 RX MEDLINE=88000549; PubMed=3307909;
 RA Dunbar J.C., Bradshaw R.A.;
 RT "Amino acid sequence of guinea pig prostate kallikrein.";
 RL Biochemistry 26:3471-3478(1987).
 CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
 CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-Xaa bonds in
 CC small molecule substrates. Highly selective action to release
 CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
 CC Met-I-Xaa or Leu-I-Xaa.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 DR PIR; A27207; A27207.
 DR HSSP; P00757; ISGF.
 DR MEROPS; S01.160; -;
 DR InterPro: IPR001314; Chymotrypsin.


```

DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS0240; TRYP_SIN_DOM; 1.
DR PROSITE: PS0134; TRYP_SIN_HIS; 1.
DR PROSITE: PS0135; TRYP_SIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein.
FT ACT_SITE 41 41 CHARGE RELAY SYSTEM.
FT ACT_SITE 96 96 CHARGE RELAY SYSTEM.
FT ACT_SITE 191 191 CHARGE RELAY SYSTEM.
FT CARBOHYD 78 78 N-LINKED (GLCNAC. . .).
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .).
FT DISULFID 7 151 BY SIMILARITY.
FT DISULFID 26 42 BY SIMILARITY.
FT DISULFID 128 197 BY SIMILARITY.
FT DISULFID 162 176 BY SIMILARITY.
FT DISULFID 187 212 BY SIMILARITY.
FT VARIANT 50 50 K -> W.
SQ SEQUENCE 239 AA; 56DC81BC10D49A64 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 239;
Best Local Similarity 66.7%; Pred. No. 4.6;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 193 GPLICDGLV 201
|||:|:|

RESULT 21
TRYX_GADMO
ID TRYX_GADMO STANDARD; PRT; 241 AA.
AC P16049; Q91040; Q92156;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Trypsin I precursor (EC 3.4.21.4).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pyloric caeca;
RA MEDLINE=94039130; PubMed=8223632;
RA Gudmundsdottir A., Gudmundsdottir E., Oskarsson S., Bjarnason J.B.,
RA Eakin A.E., Craik C.S.;
RA "Isolation and characterization of cDNAs from Atlantic cod encoding
RT two different forms of trypsinogen.";
RL Eur. J. Biochem. 217:1091-1097(1993).
RN [2]
RP SEQUENCE OF 20-58.
RC TISSUE=Pyloric caeca;
RA MEDLINE=89210867; PubMed=2707266;
RA Asgerisson B., Fox J.W., Bjarnason J.B.;
RA "Purification and characterization of trypsin from the poikilotherm
RT Gadus morhua.";
RL Eur. J. Biochem. 180:85-94(1989).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

CC EMBL: X76886; CAA54214.1;
DR PIR: S03570; S03570.
DR HSSP: P35031; 2TBS.
DR MEROPS: S01.151; -.
DR InterPro: IPR001314; Chymotrypsin.
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00722; CHYMOTRYPSIN.
DR SMART: SM00020; TRYP_SPC; 1.
DR PROSITE: PS0240; TRYP_SIN_DOM; 1.
DR PROSITE: PS0134; TRYP_SIN_HIS; 1.
DR PROSITE: PS0135; TRYP_SIN_SER; 1.
KW Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
KW Multigene family.
FT SIGNAL 1 13 POTENTIAL.
FT PROPEP 14 19 ACTIVATION PEPTIDE.
FT CHAIN 20 241 TRYPSIN I.
FT ACT_SITE 59 59 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 103 103 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 195 195 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 26 155 BY SIMILARITY.
FT DISULFID 44 60 BY SIMILARITY.
FT DISULFID 128 228 BY SIMILARITY.
FT DISULFID 135 201 BY SIMILARITY.
FT DISULFID 166 180 BY SIMILARITY.
FT DISULFID 191 215 BY SIMILARITY.
FT SITE 189 189 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
FT CONFLICT 25 25 E -> Q (IN REF. 2).
FT CONFLICT 27 28 TK -> EA (IN REF. 2).
FT CONFLICT 43 43 F -> Y (IN REF. 2).
FT CONFLICT 49 52 VSKD -> IN (IN REF. 2).
SQ SEQUENCE 241 AA; 44EC9A0106AD1A68 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 241;
Best Local Similarity 66.7%; Pred. No. 4.7;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 197 GPVVCNGVL 205
|||:|:|

RESULT 22
TRYX_GADMO
ID TRYX_GADMO STANDARD; PRT; 241 AA.
AC Q91041;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Trypsin X precursor (EC 3.4.21.4).
OS Gadus morhua (Atlantic cod).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Paracanthopterygii; Gadiformes; Gadoidei; Gadidae;
OC Gadus.
OX NCBI_TaxID=8049;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Pyloric caeca;
RA MEDLINE=94039130; PubMed=8223632;
RA Gudmundsdottir A., Gudmundsdottir E., Oskarsson S., Bjarnason J.B.,
RA Eakin A.E., Craik C.S.;
RA "Isolation and characterization of cDNAs from Atlantic cod encoding
RT two different forms of trypsinogen.";
RL Eur. J. Biochem. 217:1091-1097(1993).
CC -|- CATALYTIC ACTIVITY: Preferential cleavage: Arg-|-Xaa, Lys-|-Xaa.
CC -|- SUBCELLULAR LOCATION: Extracellular.
CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: X76887; CAA34215.1; -
HSSP; P00763; IDPO.
MEROPS; S01.151; -
InterPro: IPR001314; Chymotrypsin.
InterPro: IPR001254; Trypsin.
Pfam: PF00089; Trypsin; 1.
PRINTS; PR00722; CHYMOTRYPSIN.
SMART; SM00020; TRYP_SPC; 1.
PROSITE; PS0240; TRYP_SIN_DOM; 1.
PROSITE; PS00134; TRYP_SIN_HIS; 1.
PROSITE; PS00135; TRYP_SIN_SER; 1.
Hydrolase; Serine protease; Digestion; Pancreas; Zymogen; Signal;
Multigene family.
FT SIGNAL 1 13 POTENTIAL.
FT PROPEP 14 19 ACTIVATION PEPTIDE.
FT CHAIN 20 241 TRYPSIN X.
FT ACT_SITE 59 59 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 103 103 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 26 155 BY SIMILARITY.
FT DISULFID 44 60 BY SIMILARITY.
FT DISULFID 128 228 BY SIMILARITY.
FT DISULFID 135 201 BY SIMILARITY.
FT DISULFID 166 180 BY SIMILARITY.
FT DISULFID 191 215 BY SIMILARITY.
FT SITE 189 215 REQUIRED FOR SPECIFICITY (BY SIMILARITY).
SQ SEQUENCE 241 AA; 25976 MW; 853D7C36BCAF9DD7 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 241;
Best Local Similarity 66.7%; Pred. No. 4.7;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCGRTL 9
II:II I I
Db 197 GPVVCNVL 205

RESULT 23
KLKA_RAT STANDARD; PRT; 244 AA.
AC P36375;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glandular kallikrein 10 precursor (EC 3.4.21.35) (Tissue kallikrein)
DE (T-kininogenase) (K10) (Proteinase B) (Endopeptidase K) (Fragment).
GN KLK10 OR KLK-10.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE-Submandibular gland, and kidney;
RX MEDLINE-93041794; PubMed-1420203;
RA Ma J.-X., Chao J., Chao L.;
RT "Molecular cloning and characterization of rklk10, a cDNA encoding T-kininogenase from rat submandibular gland and kidney.";
RL Biochemistry 31:10922-10928(1992).
RN [2]
RP SEQUENCE OF 10-32; 95-124 AND 179-232.
RX MEDLINE-91224135; PubMed-2026164;
RA Gutman N., Elmoujahed A., Brillard M., du Sorbier B., Gauthier F.;
RT "Microheterogeneity of rat submaxillary gland kallikrein k10, a member of the kallikrein family.";
RT Eur. J. Biochem. 197:425-429(1991).
RN [3]

RP SEQUENCE OF 10-32 AND 97-133.
RC TISSUE-Submaxillary gland;
RX MEDLINE-90153911; PubMed-2303430;
RA Xiong W., Chen L.-M., Chao J.;
RT "Purification and characterization of a kallikrein-like T-kininogenase.";
RL J. Biol. Chem. 265:2822-2827(1990).
RN [4]
RP SEQUENCE OF 10-32 AND 97-117.
RX MEDLINE-88198057; PubMed-3482210;
RA Kato H., Nakanishi E., Enjoji K., Hayashi I., Oh-Ishi S., Iwanaga S.;
RT "Characterization of serine proteinases isolated from rat submaxillary gland: with special reference to the degradation of rat kininogens by these enzymes.";
RL J. Biochem. 102:1389-1404(1987).
CC -!- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS IN KININOGEN TO RELEASE LYS-BRADYKININ. THIS PROTEIN MAY BE INVOLVED IN THE REGULATION OF RENAL FUNCTION.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-Xaa bonds in small molecule substrates. Highly selective action to release kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of Met-I-Xaa or Leu-I-Xaa.
CC -!- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND HEAVY CHAIN LINKED BY A DISULFIDE BOND.
CC -!- TISSUE SPECIFICITY: KIDNEY AND SUBMANDIBULAR GLAND, WHERE IT IS FOUND IN THE GRANULAR CONVOLUTED TUBULE AND STRIATED DUCT CELLS. IT IS LIKELY THAT THE ENZYME IS MAINLY SYNTHESIZED IN THE GRANULAR CONVOLUTED TUBULES AND THEN TRANSFERRED TO OTHER TISSUES BY RELEASE INTO THE VASCULATURE OR INTERSTITIAL SPACE.
CC -!- PTM: PROBABLY N- AND O-GLYCOSYLATED. IT HAS CARBOHYDRATE MOIETIES OF ALPHA-METHYL-D-MANNOSIDE AND N-ACETYL-D-GLUCOSAMINE GROUPS.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL: S48142; AAB24071.1; -
PIR; A35545; A35545.
PIR; B35545; B35545.
PIR; A4284; A4284.
HSSP; P00759; ITON.
DR MEROPS; S01.165; -
DR InterPro: IPR001254; Trypsin.
DR Pfam: PF00089; trypsin; 1.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS0240; TRYP_SIN_DOM; 1.
DR PROSITE; PS00134; TRYP_SIN_HIS; 1.
DR PROSITE; PS00135; TRYP_SIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Signal; Multigene family.
FT NON_TER 1 1
FT SIGNAL <1 3 PROBABLE.
FT PROPEP 4 9 ACTIVATION PEPTIDE (PROBABLE).
FT CHAIN 10 244 GLANDULAR KALLIKREIN 10.
FT CHAIN 10 96 T-KININOGENASE LIGHT CHAIN.
FT CHAIN 97 244 T-KININOGENASE HEAVY CHAIN.
FT ACT_SITE 48 48 CHARGE RELAY SYSTEM.
FT ACT_SITE 103 103 CHARGE RELAY SYSTEM.
FT ACT_SITE 196 196 CHARGE RELAY SYSTEM.
FT DISULFID 16 156 BY SIMILARITY.
FT DISULFID 33 49 BY SIMILARITY.
FT DISULFID 135 202 BY SIMILARITY.
FT DISULFID 167 181 BY SIMILARITY.
FT DISULFID 192 217 BY SIMILARITY.
FT CARBOHYD 76 76 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 91 91 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 126 126 O-LINKED (POTENTIAL).
FT CARBOHYD 142 142 O-LINKED (POTENTIAL).

FT	DISULFID	146	218.	BY SIMILARITY.
FT	DISULFID	181	199	BY SIMILARITY.
FT	DISULFID	208	233	BY SIMILARITY.
SQ	SEQUENCE	254 AA; 27250 MW; ADA9A8A22BEFC	CRC64;	
	Query Match	74.0%;	Score 37;	DB 1; Length 254;
	Best Local Similarity	77.8%;	Pred No. 4.9;	
	Matches	7; Conservative	0; Mismatches	2; Indels
Qy	1	GPLVCRGTL	9	
Db	214	GPLVCGNTL	222	

```

RESULT 25 .
KLK4_HUMAN
ID      KLK4_HUMAN      STANDARD;          PRT;      254 AA.
AC      Q9Y5K2; Q9UBJ6; Q9GZL6;
DT      16-OCT-2001 (Rel. 40, Created)
DT      16-OCT-2001 (Rel. 40, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Kallikrein 4 precursor (EC 3.4.21.-) (Protease) (Kallikrein-like
DE      protein 1) (KLK-L1) (Enamel matrix serine proteinase 1).
DE      KLK4 OR PRSS17 OR PSTS OR EWSPL.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RP      MEDLINE=99179024; PubMed=10077646;
RX      Nelson P.S., Gan L., Ferguson C., Moss P., Gelinas R., Hood L.,
RA      Wang K.;
RA      "Molecular cloning and characterization of prostate, an androgen-
RT      regulated serine protease with prostate-restricted expression.";
RT      Proc. Natl. Acad. Sci. U.S.A. 96:3114-3119(1999).
RN      [2]
SEQUENCE FROM N.A.
RP      MEDLINE=99413477; PubMed=10485467;
RX      Yousef G.M., Obiezu C.V., Luo L.-Y., Black M.H., Diamandis E.P.;
RT      "Protease/KLK-L1 is a new member of the human kallikrein gene family,
RT      is expressed in prostate and breast tissues, and is hormonally
RT      regulated *."

```

[3]
 cancer RES. 59:4202-4250(1999)).
 RN
 RN SEQUENCE FROM N.A.
 RX MEDLINE=99367447; PubMed=10438493;
 RX MEDLINE=99367447;
 RA Stephens S.A., Verity K., Ashworth L.K., Clements J.A.;
 RT "Localization of a new prostate-specific antigen-related serine
 RT protease gene, KLK4, is evidence for an expanded human kallikrein
 RT gene family cluster on chromosome 19ql3.3-13.4";
 RL J. Biol. Chem. 274:23210-23214(1999).
 RN [4]
 RN SEQUENCE FROM N.A.
 RN MEDLINE=20510030; PubMed=11054574;
 RX Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,
 RA Moss P., Paepce B., Wang K.;
 RT "Sequencing and expression analysis of the serine protease gene
 RT cluster located in chromosome 19ql3 region.";
 RL Gene 257:119-130(2000).
 RN [5]
 RN SEQUENCE OF 22-254 FROM N.A.
 RA Simmer J.P., Ryu O.H., Qian Q., Zhang C., Cao X., Sun X., Hu C.-C.;
 RT "Cloning and characterization of a cDNA encoding human EMSPI";
 RL (in) Goldberg M. (eds.);
 RL Chemistry and biology of mineralized tissues, pp.1-1, American
 RL Academy of Orthopaedic Surgeons, Vitell (2000).
 CC -1- SUBCELLULAR LOCATION: Secreted.
 CC
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN PROSTATE.
 CC
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; AF113140; AAD21580.1; -
 DR EMBL; AF113141; AAD21581.1; -
 DR EMBL; AF135023; AAD26424.2; -
 DR EMBL; AF148532; AAD38019.1; -
 DR EMBL; AF243527; AAG33357.1; -
 DR EMBL; AF126401; AAG34246.1; -
 DR HSSP; P00763; 1DPO.
 DR MIM; 603767; -
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal.
 FT SIGNAL 1 26
 FT PROPEP 27 30
 FT CHAIN 31 254
 FT ACT_SITE 71 71
 FT ACT_SITE 116 116
 FT ACT_SITE 207 207
 FT DISULFID 37 167
 FT DISULFID 56 72
 FT DISULFID 141 241
 FT DISULFID 148 213
 FT DISULFID 178 192
 FT DISULFID 203 228
 FT CARBOHYD 169 169
 FT CONFLICT 197 197
 FT CONFLICT 197 197
 SQ SEQUENCE 254 AA; 27022 MW; 9C475E22B6EE0C8 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 254;
 Best Local Similarity 66.7%; Pred. No. 4.9;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
 |||.|||
 Db 209 GPLICNGYL 217

RESULT 26
 ID KLKF_HUMAN STANDARD; PRT; 256 AA.
 AC Q9H2R5; Q9H2R6; Q9H2R4; Q9H2R3; Q9HBG9; Q15358;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Kallikrein 15 precursor (EC 3.4.21.-) (ACO protease).
 GN KLK15
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
 RX PubMed=11010966;
 RA Yousef G.M., Scorias A., Jung K., Ashworth L.K., Diamandis E.P.;
 RT "Molecular cloning of the human kallikrein 15 gene (KLK15). Up-
 regulation in prostate cancer."
 RL J. Biol. Chem. 276:53-61(2001).
 RN [2]

RP SEQUENCE FROM N.A.
 RX MEDLINE=20510030; PubMed=11054574;
 RA Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J.,

RA Moss P., Paepfer B., Wang K.;
 RT "Sequencing and expression analysis of the serine protease gene
 RT cluster located in chromosome 19q13 region."
 RL Gene 257:119-130(2000).
 RN [3]
 RP PRELIMINARY PARTIAL SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=94289486; PubMed=8018728;
 RA Dihanich M.E., Spiess M.;
 RT "A novel serine proteinase-like sequence from human brain."
 RL Biochim. Biophys. Acta 1218:225-228(1994).
 CC -|- FUNCTION: Protease whose physiological substrate is not yet known.
 CC -|- SUBCELLULAR LOCATION: Secreted (Probable).
 CC -|- ALTERNATIVE PRODUCTS: 4 isoforms; 1 (shown here), 2, 3 and 4; are
 CC produced by alternative splicing.
 CC -|- TISSUE SPECIFICITY: Highest expression in the thyroid gland. Also
 CC expressed in the prostate, salivary, and adrenal glands and in the
 CC colon testis and kidney.
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL; AF242195; AAG09469.1; -
 DR EMBL; AF242195; AAG09470.1; -
 DR EMBL; AF242195; AAG09471.1; -
 DR EMBL; AF242195; AAG09472.1; -
 DR EMBL; AF243527; AAG33354.1; -
 DR EMBL; X75363; CAA53145.1; ALT_SEQ.
 DR MEROPS; S01.310; -
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; FALSE NEG.
 KW Hydrolase; Serine protease; Glycoprotein; Signal; Zymogen;
 KW Alternative splicing
 FT SIGNAL 1 16
 FT PROPEP 17 21
 FT CHAIN 22 256
 FT ACT_SITE 62 62
 FT ACT_SITE 106 106
 FT ACT_SITE 209 209
 FT CARBOHYD 171 171
 FT CARBOHYD 232 232
 FT VARSPPLIC 122 206
 FT VARSPPLIC 122 256
 FT VARSPPLIC 161 161
 FT VARSPPLIC 162 256
 FT CONFLICT 147 160
 FT CONFLICT 147 160
 SQ SEQUENCE 256 AA; 28087 MW; B5EBF8D602786B5 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 256;
 Best Local Similarity 77.8%; Pred. No. 5;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
 |||||
 Db 211 GPLVCGGIL 219

RESULT 27
 GRAK_RAT
 ID GRAK_RAT STANDARD; PRT; 258 AA.

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; M11565; AAA41466.1; -;
 CC EMBL; M23878; AAA42259.1; -;
 CC EMBL; M23877; AAA42259.1; JOINED.
 CC EMBL; M26533; AAA42081.1; -;
 CC PIR; A00945; KQRTN.
 CC PIR; A30971; A30971.
 CC PIR; A34050; A34050.
 CC PIR; C23863; C23863.
 CC PIR; B33359; B33359.
 CC PIR; A32340; A32340.
 CC PDB; 1TON; 16-JAN-88.
 CC MEMOPS; S01.172; -;
 CC InterPro; IPR001314; Chymotrypsin.
 CC InterPro; IPR001254; Trypsin.
 CC Pfam; PF00089; trypsin; 1.
 CC PRINTS; PR00722; CHYMOTRYPSIN.
 CC SMART; SM00020; Tryp_SPC; 1.
 CC PROSITE; PS00240; TRYPSIN_DOM; 1.
 CC PROSITE; PS00134; TRYPSIN_HIS; 1.
 CC PROSITE; PS00135; TRYPSIN_SER; 1.
 CC Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;
 KW Signal; 3D-structure.
 FT SIGNAL 1 18
 FT PROPEP 19 24
 FT CHAIN 25 259
 FT ACT_SITE 63 63
 FT ACT_SITE 118 118
 FT ACT_SITE 211 211
 FT DISULFID 31 171
 FT DISULFID 48 64
 FT DISULFID 150 217
 FT DISULFID 182 196
 FT DISULFID 207 232
 FT CARBOHYD 106 106
 FT CARBOHYD 189 189
 FT STRAND 26 26
 FT STRAND 29 30
 FT TURN 33 34
 FT TURN 37 38
 FT STRAND 39 43
 FT STRAND 47 54
 FT TURN 55 56
 FT STRAND 57 60
 FT HELIX 62 64
 FT STRAND 70 73
 FT STRAND 77 77
 FT TURN 78 79
 FT TURN 83 84
 FT STRAND 86 88
 FT STRAND 90 95
 FT TURN 97 98
 FT TURN 116 117
 FT STRAND 120 124
 FT TURN 146 147
 FT STRAND 149 154
 FT STRAND 168 168
 FT STRAND 170 177
 FT HELIX 179 181
 FT HELIX 183 186
 FT TURN 188 189
 FT HELIX 190 193
 FT STRAND 194 198
 FT TURN 200 201
 FT STRAND 205 205
 FT TURN 208 209
 FT TURN 211 212
 FT STRAND 214 217

FT TURN 218 219
 FT STRAND 220 225
 FT TURN 235 236
 FT STRAND 239 243
 FT HELIX 244 257
 SQ SEQUENCE 259 AA; 28248 MW; 3D6E60D011F926B4 CRC64;
 Query Match 74.0%; Score 37; DB 1; Length 259;
 Best Local Similarity 66.7%; Pred. No. 5;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GPLVCRGTL 9
 |||:| | |
 Db 213 GPLICDGVLT 221
 RESULT 29
 KLK9_RAT
 ID KLK9_RAT STANDARD; PRT; 259 AA.
 AC P07647;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Glandular kallikrein 9, submandibular precursor (EC 3.4.21.35)
 DE (Tissue kallikrein) (S3 kallikrein) (Submandibular enzymatic
 DE vasoconstrictor) (SEV) (KLK-S3).
 GN KLK9 OR KLK-9.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86051477; PubMed=2998455;
 RA Ashley P.L., McDonald R.J.;
 RT "Kallikrein-related mRNAs of the rat submaxillary gland: nucleotide
 RL sequences of four distinct types including tonin."; Biochemistry 24:4512-4520(1985).
 RN [2]
 RP SEQUENCE OF 25-36 AND 112-122.
 RC TISSUE=Submandibular gland;
 RX MEDLINE=91161590; PubMed=1900513;
 RA Yamaguchi T., Carretero O.A., Scicli A.G.;
 RT "A novel serine protease with vasoconstrictor activity coded by the
 RL kallikrein gene S3."; J. Biol. Chem. 266:5011-5017(1991).
 RN [3]
 RP SEQUENCE OF 25-47 AND 112-135, AND CHARACTERIZATION.
 RX MEDLINE=92250562; PubMed=1315752;
 RA Moreau T., Brillard-Bourdet M., Bouhnik J., Gauthier F.;
 RT "Protein products of the rat kallikrein gene family. Substrate
 RL specificities of kallikrein rK2 (tonin) and kallikrein rK9."; J. Biol. Chem. 267:10045-10051(1992).
 RN [4]
 RP SEQUENCE OF 25-53 AND 112-130, AND CHARACTERIZATION.
 RC TISSUE=Submandibular gland;
 RX MEDLINE=92162030; PubMed=1536657;
 RA Berg T., Schoeyen H., Wassdal I., Huli R., Gerskowitch V.P., Toft P.;
 RT "Characterization of a new kallikrein-like enzyme (KLK-S3) of the rat
 RL submandibular gland."; Biochem. J. 281:819-828(1992).
 CC -!- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
 CC IN KININOGEN TO RELEASE LYS-BRADYKININ. THIS ENZYME HAS A
 CC VASOCONSTRICTOR ACTIVITY. KLK-9 HAS BOTH A CHYMOTRYPSIN-LIKE
 CC AND A TRYPSIN-LIKE PROPERTIES.
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-xaa bonds in
 CC small molecule substrates. Highly selective action to release
 CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
 CC Met-I-Xaa or Leu-I-Xaa.
 CC -!- SUBUNIT: HETERODIMER OF A LIGHT CHAIN AND HEAVY CHAIN LINKED
 CC BY A DISULFIDE BOND.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M11566; AAA41467.1; -
CC DR PIR; D23863; D23863.
CC DR HSSP; P00759; ITON.
CC DR MEROPS; S01.407; -
CC DR InterPro; IPR001314; Chymotrypsin.
CC DR InterPro; IPR001254; Trypsin.
CC DR Pfam; PF00089; trypsin; 1.
CC DR PRINTS; PR00722; CHYMOTRYPSIN.
CC DR SMART; SM00020; TRYPSIN_SPC; 1.
CC DR PROSITE; PS0240; TRYPSIN_DOM; 1.
CC DR PROSITE; PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE; PS00135; TRYPSIN_SER; 1.
CC DR Hydrolase; Serine protease; Multigene family; Zymogen;
CC KW Signal.
CC KW
CC FT SIGNAL 1 18 PROBABLE.
CC FT PROPEP 19 24 ACTIVATION PEPTIDE.
CC FT CHAIN 25 259 GLANDULAR KALLIKREIN 9.
CC FT CHAIN 25 111 LIGHT CHAIN.
CC FT CHAIN 112 259 HEAVY CHAIN.
CC FT ACT_SITE 63 63 CHARGE RELAY SYSTEM.
CC FT ACT_SITE 118 118 CHARGE RELAY SYSTEM.
CC FT ACT_SITE 211 211 CHARGE RELAY SYSTEM.
CC FT DISULFID 31 171 BY SIMILARITY.
CC FT DISULFID 48 64 BY SIMILARITY.
CC FT DISULFID 150 217 BY SIMILARITY.
CC FT DISULFID 182 196 BY SIMILARITY.
CC FT DISULFID 207 232 BY SIMILARITY.
CC FT CARBOHYD 106 106 N-LINKED (GLCNAC...) (PROBABLE).
CC SQ SEQUENCE 259 AA; 28367 MW; D167E8518BEC0791 CRC64;
Query Match 74.0%; Score 37; DB 1; Length 259;
Best Local Similarity 66.7%; Pred. No. 5;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCGRTL 9
|||:|:|
Db 213 GPLICDGLV 221
RESULT 30
KLKM_MOUSE STANDARD; PRT; 259 AA.
AC P15948;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glandular kallikrein K22 precursor (EC 3.4.21.35) (Tissue kallikrein)
DE (MGK-22) (Epidermal growth factor-binding protein type A) (EGF-BP A)
DE (Nerve growth factor beta chain endopeptidase) (Beta-NGF-
DE endopeptidase).
GN KLK22 OR KLK-22.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RN SEQUENCE FROM N.A.
RP STRAIN=BALB/C; TISSUE=Salivary gland;
RC MEDLINE=88107594; PubMed=332387;
RX Drinkwater C.C., Evans B.A., Richards R.I.;
RA "Mouse glandular kallikrein genes: identification and
RT characterization of the genes encoding the epidermal growth factor
RT binding proteins.";
RL Biochemistry 26:6750-6756(1987).

RN (2)
RP SEQUENCE OF 25-54.
RX MEDLINE=91190897; PubMed=2012805;
RA Fahnestock M., Woo J.E., Lopez G.A., Snow J., Walz D.A., Arici M.J.,
RA Mobley W.C.;
RT "beta-NGF-endopeptidase: structure and activity of a kallikrein
RT encoded by the gene mgk-22.";
RL Biochemistry 30:3443-3450(1991).
RN (3)
RP SEQUENCE OF 25-41.
RC TISSUE=Submandibular gland;
RX MEDLINE=92348355; PubMed=1639762;
RA Peters J., Takahashi S., Tada M., Miyake Y.;
RT "mgk-6-derived true tissue kallikrein is synthesized, and
RT targeted through a regulated secretory pathway in mouse pituitary
RT AtT-20 cells.";
RL J. Biochem. 111:643-648(1992).
RN (4)
RP SEQUENCE OF 17-54 AND 70-120 FROM N.A.
RX MEDLINE=87250386; PubMed=3036794;
RA Evans B.A., Drinkwater C.C., Richards R.I.;
RT "Mouse glandular kallikrein genes. Structure and partial sequence
RT analysis of the kallikrein gene locus.";
RL J. Biol. Chem. 262:8027-8034(1987).
CC -!- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -!- CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-Xaa bonds in
CC small molecule substrates. Highly selective action to release
CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
CC Met-I-Xaa or Leu-I-Xaa.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M17979; AAA37682.1; ALT_SEQ.
CC DR EMBL; M17977; AAA37682.1; JOINED.
CC DR EMBL; M17978; AAA37682.1; JOINED.
CC DR EMBL; M18598; AAA39361.1; -
CC DR EMBL; M18618; AAA39362.1; -
CC DR PIR; A29746; A29746.
CC DR PIR; A38356; A38356.
CC DR HSSP; P00757; LSGF.
CC DR MEROPS; S01.039; -
CC DR MGD; MGI:95291; K1K22.
CC DR InterPro; IPR001314; Chymotrypsin.
CC DR InterPro; IPR001254; Trypsin.
CC DR Pfam; PF00089; trypsin; 1.
CC DR PRINTS; PR00722; CHYMOTRYPSIN.
CC DR SMART; SM00020; TRYPSIN_SPC; 1.
CC DR PROSITE; PS0240; TRYPSIN_DOM; 1.
CC DR PROSITE; PS00134; TRYPSIN_HIS; 1.
CC DR PROSITE; PS00135; TRYPSIN_SER; 1.
CC KW Hydrolase; Serine protease; Multigene family; Zymogen;
CC KW Signal.
CC FT SIGNAL 1 17
CC FT PROPEP 18 24 ACTIVATION PEPTIDE.
CC FT CHAIN 25 259 GLANDULAR KALLIKREIN K22.
CC FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
CC FT ACT_SITE 118 118 CHARGE RELAY SYSTEM.
CC FT ACT_SITE 211 211 CHARGE RELAY SYSTEM.
CC FT DISULFID 31 171 BY SIMILARITY.
CC FT DISULFID 50 66 BY SIMILARITY.
CC FT DISULFID 150 217 BY SIMILARITY.
CC FT DISULFID 182 196 BY SIMILARITY.
CC FT DISULFID 207 232 BY SIMILARITY.
CC FT CARBOHYD 102 102 N-LINKED (GLCNAC...) (PROBABLE).

```
SQ SEQUENCE 259 AA; 28384 MW; D7745794D8A87B9C CRC64;
Query Match 74.0%; Score 37; DB 1; Length 259;
Best Local Similarity 66.7%; Pred. No. 5;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
   |||:| | |
Db 213 GPLICDGL 221

RESULT 31
NRPN_MOUSE
ID NRPN_MOUSE STANDARD; PRT; 260 AA.
AC Q61955;
DT 15-JUL-1999 (Rel. 38, Created)
DT 13-JUL-1999 (Rel. 38, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Neuropilin precursor (EC 3.4.21.-) (NP) (Kallikrein 8).
GN K1K8 OR PRSS19 OR NRPN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-Hippocampus;
RX MEDLINE=95348817; PubMed=7623137;
RA Chen Z.-L., Yoshida S., Kato K., Momota Y., Suzuki J., Tanaka T.,
RA Ito J., Nishino H., Aimoto S., Kiyama H., Shiosaka S.;
RT "Expression and activity-dependent changes of a novel limbic-serine
RT protease gene in the hippocampus.";
RL J. Neurosci. 15:5088-5097(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Yoshida S., Hirata A., Inoue N., Shiosaka S.;
RT "Cloning and assignment of mouse neuropilin gene, Prss19 to chromosome
RT 7B4.";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF N-TERMINUS, AND CHARACTERIZATION.
RC STRAIN-BALB/C; TISSUE-Brain;
RX MEDLINE=98225202; PubMed=9556608;
RA Shimizu C., Yoshida S., Shibata M., Kato K., Momota Y., Matsumoto K.,
RA Shiosaka T., Midorikawa R., Kamachi T., Kawabe A., Shiosaka S.;
RT "Characterization of recombinant and brain neuropilin, a
RT plasticity-related serine protease.";
RL J. Biol. Chem. 273:11189-11196(1998).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.1 ANGSTROMS) OF 33-257.
RC TISSUE-Hippocampus;
RX MEDLINE=99134351; PubMed=9933620;
RA Kishi T., Kato M., Shimizu T., Kato K., Matsumoto K., Yoshida S.,
RA Shiosaka S., Hakoshima T.;
RT "Crystal structure of neuropilin, a hippocampal protease involved in
RT kindling epileptogenesis.";
RL J. Biol. Chem. 274:4220-4224(1999).
CC -1- FUNCTION: SUGGESTED TO BE INVOLVED IN KINDLING EPILEPTOGENESIS AND
CC HIPPOCAMPAL PLASTICITY. HAS A STRONG PROTEOLYTIC ACTIVITY AGAINST
CC FIBRONECTIN.
CC -1- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: ARG-, LYS-.
CC -1- ENZYME REGULATION: STRONGLY INHIBITED BY DIISOPROPYL
CC FLUOROPHOSPHATE, LEUPEPTIN AND (4-AMIDINOPHENYL)METHANESULFONYL 1-
CC FLUORIDE.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- TISSUE SPECIFICITY: EXPRESSED SPECIFICALLY IN THE LIMBIC SYSTEM OF
CC MOUSE BRAIN AND IS LOCALIZED AT HIGHEST CONCENTRATION IN PYRAMIDAL
CC NEURONS OF THE HIPPOCAMPAL CA1-3 SUBFIELDS.
CC -1- MASS SPECTROMETRY: MW=26613; METHOD=MALDI; RANGE=29-260.
CC -1- MASS SPECTROMETRY: MW=26229; METHOD=MALDI; RANGE=33-260.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
ENBL; D30785; BAA06451.1; -.
EMBL; AB032202; BAA92435.1; -.
PDB; INPM; 23-MAR-99.
DR MEROPS; S01.244; -.
DR MGD; MGI:892018; K1k8.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PRO0722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS0240; TRYP_SIN_DOM; 1.
DR PROSITE; PS00134; TRYP_SIN_HIS; 1.
DR PROSITE; PS00135; TRYP_SIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
KW 3D-structure.
FT SIGNAL 1 28 POTENTIAL.
FT PROPEP 29 32
FT CHAIN 33 260 NEUROPSIN.
FT ACT_SITE 73 73 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 212 212 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 39 173
FT DISULFID 58 74
FT DISULFID 145 246
FT DISULFID 152 218
FT DISULFID 184 198
FT DISULFID 208 233
FT CARBOHYD 110 110 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 260 AA; 28523 MW; BE5F6F6BE37CD60E CRC64;

Query Match 74.0%; Score 37; DB 1; Length 260;
Best Local Similarity 77.8%; Pred. No. 5;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
   ||| | | |
Db 214 GPLVCDGML 222

RESULT 32
KLK1_MOUSE
ID KLK1_MOUSE STANDARD; PRT; 261 AA.
AC P15947; O61855.
DT 01-APR-1990 (Rel. 14, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Glandular kallikrein K1 precursor (EC 3.4.21.35) (Tissue kallikrein)
DE (MGK-6) (Renal kallikrein) (KAL-B).
GN KLK1 OR KLK-1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=86168299; PubMed=3007510;
RA van Leeuwen B.H., Evans B.A., Tregear G.W., Richards R.I.;
RT "Mouse glandular kallikrein genes. Identification, structure, and
RT expression of the renal kallikrein gene.";
RL J. Biol. Chem. 261:5529-5535(1986).
RN [2]
RP SEQUENCE FROM N.A.
RA Tada M., Peters J., Takahashi S., Inoue H., Miyake Y.;
RT "Identification of a tissue kallikrein gene, mGK-6, expressed in a
RT mouse neuroendocrine cell line.";
```


Submitted (JUN-1991) to the EMBL/GenBank/DBJ databases.

```

RL RN SEQUENCE FROM N.A.
RP RP STRAIN=C57BL/6J; TISSUP=Kidney;
RX RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukuishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaudo I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staebli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
[4]
SEQUENCE OF 25-44.
RP TISSUP=Submandibular gland;
RX MEDLINE=92348355; PubMed=1639762;
RA Peters J., Takahashi S., Tada M., Miyake Y.;
RT "mCK-6-derived true tissue kallikrein is synthesized, processed, and
RT targeted through a regulated secretory pathway in mouse pituitary
RT ACT-20 cells.";
RL J. Biochem. 111:643-648(1992).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-Xaa bonds in
CC small molecule substrates. Highly selective action to release
CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
CC Met-I-Xaa or Leu-I-Xaa.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL; M13500; AAG11389.1; -.
DR EMBL; M13498; AAG11389.1; JOINED.
DR EMBL; M13499; AAG11389.1; JOINED.
DR EMBL; D10464; BAA01257.1; -.
DR EMBL; AK002278; BAB21982.1; -.
DR PIR; A25606; A25606.
DR HSSP; P00757; ISGF.
DR MEROPS; S01.160; -.
DR MGD; MGI:102850; Kikl.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS00240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
DR Hydrolase; Serine protease; Multigene family; Zymogen;
KW Signal.
FT SIGNAL 1 18 PROBABLE.
FT PROPEP 19 24 ACTIVATION PEPTIDE.
FT CHAIN 25 261 GLANDULAR KALLIKREIN K1.

```

```

FT ACT_SITE 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 CHARGE RELAY SYSTEM.
FT DISULFID 31 BY SIMILARITY.
FT DISULFID 50 BY SIMILARITY.
FT DISULFID 152 BY SIMILARITY.
FT DISULFID 184 BY SIMILARITY.
FT DISULFID 209 BY SIMILARITY.
FT CARBOHYD 102 N-LINKED (GLCNAC... ) (PROBABLE).
FT CONFLICT 57 A -> V (IN REF. 1).
SQ SEQUENCE 261 AA; 7875 DDFFBFF94B8 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 5;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 215 GPLICDGV 223

RESULT 33
KLK3_MOUSE
ID KLK3_MOUSE STANDARD; PRT; 261 AA.
AC P00756;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glandular kallikrein K3 precursor (BC 3.4.21.35) (Tissue kallikrein)
DE (mCK-3) (7S nerve growth factor gamma chain) (Gamma-NGF).
OS KLK3 OR KLK-3 OR NGFG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85076169; PubMed=6548955;
RA Ullrich A., Gray A., Wood W.I., Hayflick J., Seeburg P.H.;
RT "Isolation of a cDNA clone coding for the gamma-subunit of mouse
RT nerve growth factor using a high-stringency selection procedure.";
RL DNA 3:387-392(1984).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=85257431; PubMed=3848399;
RA Evans B.A., Richards R.I.;
RT "Genes for the alpha and gamma subunits of mouse nerve growth factor
RT are contiguous.";
RL EMBO J. 4:133-138(1985).
RN [3]
RP SEQUENCE OF 25-261.
RX MEDLINE=81264363; PubMed=7263706;
RA Thomas K.A., Baglan N.C., Bradshaw R.A.;
RT "The amino acid sequence of the gamma-subunit of mouse submaxillary
RT gland 7 S nerve growth factor.";
RL J. Biol. Chem. 256:9156-9166(1981).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (3.15 ANGSTROMS) OF 7S COMPLEX.
RC STRAIN=SWISS WEBSTER; TISSUE=Submaxillary gland;
RX MEDLINE=98035451; PubMed=9351801;
RA Bax B., Blundell T.L., Murray-Rust J., McDonald N.Q.;
RT "Structure of mouse 7S NGF: a complex of nerve growth factor with
RT four binding proteins.";
RL Structure 5:1275-1285(1997).
CC -1- FUNCTION: 7S NGF ALPHA CHAIN STABILIZES THE 7S COMPLEX. THE BETA
CC DIMER PROMOTES NEURITE GROWTH. THE GAMMA CHAIN IS AN ARGININE-
CC SPECIFIC PROTEASE; IT MAY ALSO HAVE PLASMINOGEN ACTIVATOR
CC ACTIVITY, AS WELL AS MITOGENIC ACTIVITY FOR CHICK EMBRYO
CC FIBROBLASTS.
CC -1- SUBUNIT: 7S NERVE GROWTH FACTOR IS COMPOSED OF TWO ALPHA CHAINS,
CC A BETA DIMER COMPOSED OF IDENTICAL CHAINS, AND TWO GAMMA CHAINS.
CC -1- MISCELLANEOUS: THIS PRECURSOR IS CLEAVED INTO SEGMENTS TO PRODUCE
CC THE ACTIVE FORM OF THE GAMMA CHAIN, WHICH OCCURS NATURALLY AS

```

CC COMBINATIONS OF EITHER TWO OR THREE SEGMENTS HELD TOGETHER BY
 CC DISULFIDE BONDS: B1 + A OR B1 + C + B2.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X01389; CAA25645.1; -;
 CC EMBL: X01798; CAA25928.1; -;
 CC EMBL: X01799; CAA25930.1; -;
 CC PIR: A00942; NGSNG
 CC PDB: 1SGF; 27-MAY-98.
 CC MEROPS: S01.170; -;
 CC MGD: MGI:97322; Ngfg.
 CC InterPro: IPR001314; Chymotrypsin.
 CC Pfam: PF00089; trypsin; 1.
 CC PRINTS: PR00722; Chymotrypsin.
 CC SMART: SM00020; Tryp_SPC; 1.
 CC PROSITE: PS00240; TRYPSIN_DOM; 1.
 CC PROSITE: PS00134; TRYPSIN_HIS; 1.
 CC PROSITE: PS00135; TRYPSIN_SER; 1.
 CC Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;
 CC Signal; Growth factor; 3D-structure.
 CC FT SIGNAL 1 18 PROBABLE.
 CC FT PROPEP 19 24 ACTIVATION PEPTIDE.
 CC FT CHAIN 25 261 GLANDULAR KALLIKREIN K3.
 CC FT CHAIN 25 107 NERVE GROWTH FACTOR GAMMA CHAIN 1.
 CC FT CHAIN 112 261 NERVE GROWTH FACTOR GAMMA CHAIN 2.
 CC FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
 CC FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
 CC FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
 CC FT DISULFID 31 173 CHARGE RELAY SYSTEM.
 CC FT DISULFID 50 66 N-LINKED (GLCNAC...).
 CC FT DISULFID 152 219 SEGMENT B1.
 CC FT DISULFID 184 198 SEGMENT A.
 CC FT DISULFID 209 234 SEGMENT C.
 CC FT DOMAIN 25 107 SEGMENT B2.
 CC FT DOMAIN 112 261 MISSING (IN REF. 2).
 CC FT DOMAIN 165 261 MISSING (IN REF. 2).
 CC FT CONFLICT 108 111
 CC FT SEQUENCE 261 AA; 28998 MW; 4870748E174AF7C8 CRC64;
 CC -----
 CC Query Match 74.0%; Score 37; DB 1; Length 261;
 CC Best Local Similarity 66.7%; Pred. No. 5;
 CC Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 CC -----
 CC QY 1 GPLVCRGTL 9
 CC |||:| | |
 CC Db 215 GPLICDGV 223
 CC -----
 CC RESULT 34
 CC ID KLK5_MOUSE STANDARD; PRT; 261 AA.
 CC AC P15945;
 CC DT 01-APR-1990 (Rel. 14, Created)
 CC DT 01-APR-1990 (Rel. 14, Last sequence update)
 CC DT 30-MAY-2000 (Rel. 39, Last annotation update)
 CC DE Glandular kallikrein K5 precursor (EC 3.4.21.35) (Tissue kallikrein)
 CC DE (MGK-5).
 CC GN KLK5 OR KLK-5.
 CC OS Mus musculus (Mouse).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC OC Mammalia; Theria; Rodentia; Sciurognathi; Muridae; Mus.

OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BALB/C; TISSUE=Liver;
 RX MEDLINE=88096499; PubMed=3502721;
 RA Drinkwater C.C.; Richards R.I.;
 RT "Sequence of the mouse glandular kallikrein gene, mGK-5.";
 RL Nucleic Acids Res. 15:10052-10052(1987).
 RN [2]
 RP SEQUENCE OF 70-122 FROM N.A.
 RX MEDLINE=87250386; PubMed=3036794;
 RA Evans B.A.; Drinkwater C.C.; Richards R.I.;
 RT "Mouse glandular kallikrein genes. Structure and partial sequence
 analysis of the kallikrein gene locus.";
 RL J. Biol. Chem. 262:8027-8034(1987).
 CC -!- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
 CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
 CC -!- CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-Xaa bonds in
 CC small molecule substrates. Highly selective action to release
 CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
 CC Met-I-Xaa or Leu-I-Xaa.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X00500; CAA68553.1; -;
 CC EMBL: M18604; AAD3284.1; -;
 CC PIR: S06305; TRMSW5.
 CC HSSP: P00757; 1SGF.
 CC MEROPS: S01.037; -;
 CC MGD: MGI:892020; Klk5.
 CC InterPro: IPR001314; Chymotrypsin.
 CC InterPro: IPR001254; Trypsin.
 CC Pfam: PF00089; trypsin; 1.
 CC PRINTS: PR00722; Chymotrypsin.
 CC SMART: SM00020; Tryp_SPC; 1.
 CC PROSITE: PS00240; TRYPSIN_DOM; 1.
 CC PROSITE: PS00134; TRYPSIN_HIS; 1.
 CC PROSITE: PS00135; TRYPSIN_SER; 1.
 CC Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;
 CC Signal.
 CC FT SIGNAL 1 18 PROBABLE.
 CC FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
 CC FT CHAIN 25 261 GLANDULAR KALLIKREIN K5.
 CC FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
 CC FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
 CC FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
 CC FT DISULFID 31 173 BY SIMILARITY.
 CC FT DISULFID 50 66 BY SIMILARITY.
 CC FT DISULFID 152 219 BY SIMILARITY.
 CC FT DISULFID 184 198 BY SIMILARITY.
 CC FT DISULFID 209 234 BY SIMILARITY.
 CC FT CARBOHYD 102 102 N-LINKED (GLCNAC... (PROBABLE).
 CC FT SEQUENCE 261 AA; 28748 MW; 29B4D66935522D8 CRC64;
 CC -----
 CC Query Match 74.0%; Score 37; DB 1; Length 261;
 CC Best Local Similarity 66.7%; Pred. No. 5;
 CC Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 CC -----
 CC QY 1 GPLVCRGTL 9
 CC |||:| | |
 CC Db 215 GPLICDGV 223
 CC -----
 CC RESULT 35
 CC KLK6_MOUSE

ID KLK6_MOUSE STANDARD; PRT; 261 AA.
 AC P00755;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Glandular kallikrein K6 precursor (EC 3.4.21.35) (Tissue kallikrein)
 DE (MGK-1).
 GN KLK6 OR KLK-6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP STRAIN=OUAKENBUSH INBRED;
 RX MEDLINE=83219214; PubMed=6602295;
 RT Hason A.J., Evans B.A., Cox D.R., Shine J., Richards R.I.;
 RA "Structure of mouse kallikrein gene family suggests a role in
 RT specific processing of biologically active peptides.";
 RL Nature 303:300-307(1983).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=87250386; PubMed=3036794;
 RX Evans B.A., Drinkwater C.C., Richards R.I.;
 RA "Mouse glandular kallikrein genes. Structure and partial sequence
 RT analysis of the kallikrein gene locus.";
 RL J. Biol. Chem. 262:8027-8034(1987).
 CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
 CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage of Arg-|-Xaa bonds in
 CC small molecule substrates. Highly selective action to release
 CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
 CC Met-|-Xaa or Leu-|-Xaa.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; V00829; CAA24213.1; -
 DR EMBL; J00390; AAA39349.1; -
 DR PIR; A00941; KMSL.
 DR HSSP; P00757; ISGF.
 DR MEROPS; S01.164; -.
 DR MGD; MGI:892019; Klk6.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;
 KW Signal.
 FT SIGNAL 1 18 PROBABLE.
 FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
 FT CHAIN 25 261 GLANDULAR KALLIKREIN K6.
 FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
 FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
 FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
 FT DISULFID 31 173 BY SIMILARITY.
 FT DISULFID 50 66 BY SIMILARITY.
 FT DISULFID 152 219 BY SIMILARITY.
 FT DISULFID 184 198 BY SIMILARITY.
 FT DISULFID 209 234 BY SIMILARITY.
 FT CARBOHYD 102 102 N-LINKED (GLCNAC...) (PROBABLE).
 SQ SEQUENCE 261 AA; 29021 MW; 15DC946600AF1604 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 261;
 Best Local Similarity 66.7%; Pred. No. 5;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GPLVCRGTL 9
 Db 215 GPLICDGV 223
 |||:| | |
 |||:| | |
 RESULT 36
 KLK8_RAT
 ID KLK8_RAT STANDARD; PRT; 261 AA.
 AC P36374;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 01-JUN-1994 (Rel. 29, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Glandular kallikrein 8, prostatic precursor (EC 3.4.21.35) (Tissue
 DE kallikrein) (P1 kallikrein) (RGR-8).
 GN KLK8 OR KLK-8.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=89352606; PubMed=2765531;
 RX Brady J.M., Wines D.R., Macdonald R.J.;
 RA "Expression of two kallikrein gene family members in the rat
 RT prostate.";
 RL Biochemistry 28:5203-5210(1989).
 CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
 CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
 CC -1- CATALYTIC ACTIVITY: Preferential cleavage of Arg-|-Xaa bonds in
 CC small molecule substrates. Highly selective action to release
 CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
 CC Met-|-Xaa or Leu-|-Xaa.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M27217; AAA42036.1; -
 DR EMBL; M27215; AAA42036.1; JOINED.
 DR EMBL; M27216; AAA42036.1; JOINED.
 DR PIR; A34079; A34079.
 DR HSSP; P00759; ITON.
 DR MEROPS; S01.160; -.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; TRYP_SPC; 1.
 DR PROSITE; PS50240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;
 KW Signal.
 FT SIGNAL 1 18 PROBABLE.
 FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
 FT CHAIN 25 261 GLANDULAR KALLIKREIN 8, PROSTATIC.
 FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
 FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
 FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
 FT DISULFID 31 173 BY SIMILARITY.
 FT DISULFID 50 66 BY SIMILARITY.
 FT DISULFID 152 219 BY SIMILARITY.

```
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 209 234 BY SIMILARITY.
FT CARBOHYD 108 108 N-LINKED (GLCNAC... ) (POTENTIAL).
SQ SEQUENCE 261 AA; 29013 MW; CA8F12151B04E337 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 5;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 215 GPLICDGL 223

RESULT 37
KLK9_MOUSE STANDARD; PRT; 261 AA.
AC PL5949;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glandular kallikrein K9 precursor (EC 3.4.21.35) (Tissue kallikrein)
DE (MGK-9) (Epidermal growth factor-binding protein type C) (EGF-BP C).
GN KLK9 OR KLK-9 OR EGFBP3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=BALB/C; TISSUE=Salivary gland;
RX MEDLINE=88107594; PubMed=3322387;
RA Drinkwater C.C., Evans B.A., Richards R.I.;
RT "Mouse glandular kallikrein genes: identification and
RT characterization of the genes encoding the epidermal growth factor
RT binding proteins.";
RL Biochemistry 26:6750-6756(1987).
[2]
RN SEQUENCE FROM N.A.
RP MEDLINE=88107593; PubMed=3322386;
RX Blaber M., Isackson P.J., Bradshaw R.A.;
RA "A complete cDNA sequence for the major epidermal growth factor
RT binding protein in the male mouse submandibular gland.";
RL Biochemistry 26:6742-6749(1987).
[3]
RN SEQUENCE OF 16-54 AND 70-122 FROM N.A.
RP STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=87250386; PubMed=3036794;
RA Evans B.A., Drinkwater C.C., Richards R.I.;
RT "Mouse glandular kallikrein genes. Structure and partial sequence
RT analysis of the kallikrein gene locus.";
RL J. Biol. Chem. 262:8027-8034(1987).
CC -1- FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -1- CATALYTIC ACTIVITY: Preferential cleavage of Arg-|-Xaa bonds in
CC small molecule substrates. Highly selective action to release
CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
CC Met-|-Xaa or Leu-|-Xaa.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; M17962; AAA37541.1;
CC EMBL; M17985; AAA37681.1;
CC EMBL; M17983; AAA37681.1; JOINED.
CC EMBL; M17984; AAA37681.1; JOINED.
```

```
DR EMBL; M18588; AAA39350.1; ALT_SEQ.
DR EMBL; M18608; AAA39351.1; -.
DR PIR; A29745; A29745.
DR PIR; C29746; C29746.
DR HSSP; P00737; ISGF.
DR MEROPS; S01.169; -.
DR MGD; MGI:95293; K1K9.
DR InterPro; IPR001314; Chymotrypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYPSIN_DOM; 1.
DR PROSITE; PS0240; TRYPSIN_HIS; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;
KW SIGNAL.
FT SIGNAL 1 18 PROBABLE.
FT PROPEP 19 24 ACTIVATION PEPTIDE.
FT CHAIN 25 261 GLANDULAR KALLIKREIN K9.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
FT DISULFID 31 173 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 152 219 BY SIMILARITY.
FT DISULFID 184 198 BY SIMILARITY.
FT DISULFID 209 234 BY SIMILARITY.
FT CARBOHYD 102 102 N-LINKED (GLCNAC... ) (PROBABLE).
SQ SEQUENCE 261 AA; 28900 MW; 796FD4DAE56020D7 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 5;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 215 GPLICDGL 223

RESULT 38
KLKB_MOUSE STANDARD; PRT; 261 AA.
AC PL5946;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Glandular kallikrein K11 precursor (EC 3.4.21.35) (Tissue kallikrein)
DE (MGK-11).
GN KLK11 OR KLK-11.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=BALB/C; TISSUE=Liver;
RX MEDLINE=89083511; PubMed=3205728;
RA Drinkwater C.C., Richards R.I.;
RT "Sequence of mGK-11, a mouse glandular kallikrein gene.";
RL Nucleic Acids Res. 16:10918-10918(1988).
[2]
RN SEQUENCE FROM N.A.
RX MEDLINE=85257431; PubMed=3848399;
RA Evans B.A., Richards R.I.;
RT "Genes for the alpha and gamma subunits of mouse nerve growth factor
RT are contiguous.";
RL EMBO J. 4:133-138(1985).
[3]
RN SEQUENCE OF 16-54 AND 69-122 FROM N.A.
RX MEDLINE=87250386; PubMed=3036794;
RA Evans B.A., Drinkwater C.C., Richards R.I.;
RT "Mouse glandular kallikrein genes. Structure and partial sequence
```

RT analysis of the kallikrein gene locus.";
 RL J. Biol. Chem. 262:8027-8034(1987).
 CC -|- FUNCTION: GLANDULAR KALLIKREIN CLEAVE MET-LYS AND ARG-SER BONDS
 CC IN KININOGEN TO RELEASE LYS-BRADYKININ
 CC -|- CATALYTIC ACTIVITY: Preferential cleavage of Arg-|-Xaa bonds in
 CC small molecule substrates. Highly selective action to release
 CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
 CC Met-|-Xaa or Leu-|-Xaa.
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X13215; CAA31604.3; -;
 DR EMBL; X13216; CAA31604.3; JOINED.
 DR EMBL; X13217; CAA31604.3; JOINED.
 DR EMBL; X13218; CAA31604.3; JOINED.
 DR EMBL; M18590; AAA39352.1; ALT_SEQ.
 DR EMBL; M18610; AAA39353.1; -;
 DR PIR; S01971; S01971.
 DR HSP; P00757; 1SGF.
 DR MEROPS; S01.041; -;
 DR MGD; MGI:892023; Kik11.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Multigene family; Zymogen;
 KW Signal.
 FT SIGNAL 1 18 PROBABLE.
 FT PROPEP 19 24 ACTIVATION PEPTIDE.
 FT CHAIN 25 261 GLANDULAR KALLIKREIN K11.
 FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
 FT ACT_SITE 120 120 CHARGE RELAY SYSTEM.
 FT ACT_SITE 213 213 CHARGE RELAY SYSTEM.
 FT DISULFID 31 173 BY SIMILARITY.
 FT DISULFID 50 66 BY SIMILARITY.
 FT DISULFID 152 219 BY SIMILARITY.
 FT DISULFID 184 198 BY SIMILARITY.
 FT DISULFID 209 234 BY SIMILARITY.
 FT CARBOHYD 102 102 N-LINKED (GLCNAC...) (PROBABLE).
 SQ SEQUENCE 261 AA; 28727 MW; 8D32EF8D835EA7B CRC64;
 Query Match 74.0%; Score 37; DB 1; Length 261;
 Best Local Similarity 66.7%; Pred. No. 5;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
 QY 1 GPLVCRGTL 9
 DB 215 GPLICDGV 223
 RESULT 39
 KLKG_MOUSE STANDARD; PRT; 261 AA.
 AC P04071;
 DT 01-NOV-1986 (Rel. 03, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Gamma-
 GN Gamma-
 OS Mus musculus (Mouse).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RX SEQUENCE FROM N.A.
 RP MEDLINE=88243703; PubMed=3288617;
 RA Drinkwater C.C., Evans B.A., Richards R.I.;
 RT "Sequence and expression of mouse gamma-
 RL J. Biol. Chem. 263:8565-8568(1988).
 RN [2]
 RP SEQUENCE OF 16-54 AND 70-122 FROM N.A.
 RX MEDLINE=87250386; PubMed=3036794;
 RA Evans B.A., Drinkwater C.C., Richards R.I.;
 RT "Mouse glandular kallikrein genes. Structure and partial sequence
 RT analysis of the kallikrein gene locus.";
 RL J. Biol. Chem. 262:8027-8034(1987).
 RN [3]
 RP SEQUENCE OF 25-64 AND 165-184.
 RC STRAIN=SWISS WEBSTER;
 RX MEDLINE=83109049; PubMed=6337154;
 RA Poe M., Wu J.K., Florence J.R., Rodkey J.A., Bennett C.D.,
 RA Hoogsteen K.;
 RT "Purification and properties of renin and gamma-
 RT submaxillary gland.";
 RL J. Biol. Chem. 258:2209-2216(1983).
 CC -|- CATALYTIC ACTIVITY: Cleaves Leu-|-Leu bond in synthetic
 CC tetradecapeptide renin substrate, producing angiotensin I, but not
 CC active on natural angiotensinogen. Also hydrolyzes BZ-Arg-para-
 CC nitroanilide.
 CC -|- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
 CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; J03877; AAA40049.1; -;
 DR EMBL; M18594; AAA39357.1; -;
 DR EMBL; M18615; AAA39358.1; -;
 DR PIR; A28062; A28062.
 DR PIR; A05324; A05324.
 DR HSP; P36368; IAO5.
 DR MEROPS; S01.163; -;
 DR MGD; MGI:891982; Kik16.
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001254; Trypsin.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR SMART; SM00020; Tryp_SPC; 1.
 DR PROSITE; PS00240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;
 KW Signal.
 FT SIGNAL 1 18 PROBABLE.
 FT PROPEP 19 24 ACTIVATION PEPTIDE.
 FT CHAIN 25 261 GAMMA-RENIN, SUBMANDIBULAR GLAND.
 FT ACT_SITE 65 65 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
 FT DISULFID 31 173 BY SIMILARITY.
 FT DISULFID 50 66 BY SIMILARITY.
 FT DISULFID 152 219 BY SIMILARITY.
 FT DISULFID 184 198 BY SIMILARITY.
 FT DISULFID 209 234 BY SIMILARITY.
 FT CARBOHYD 102 102 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CONFLICT 45 45 H -> A (IN REF. 3).
 FT CONFLICT 48 49 HI -> YL (IN REF. 3).
 SQ SEQUENCE 261 AA; 28722 MW; 53D8984BF41E0F3D CRC64;

Query Match 74.0%; Score 37; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 5;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
DB 215 GPLICDGLV 223

RESULT 40

GRK_MOUSE STANDARD; PRT; 263 AA.
AC O35205;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Granzyme K precursor (EC 3.4.21.-).
GN GZMK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RX MEDLINE=97390469; PubMed=9242702;
RA Shresta S., Goda P., Wesselschmidt R., Ley T.J.;
RT "Residual cytotoxicity and granzyme K expression in granzyme
A-deficient cytotoxic lymphocytes";
RL J. Biol. Chem. 272:20236-20244(1997).
CC -! SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES.
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY. GRANZYME SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; AF011446; AAC17930.1; -
DR HSSP; P20160; 1A7S.
DR MGD; MGI:1298232; GzmK.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 21
FT PROPEP 22 25 ACTIVATION PEPTIDE.
FT CHAIN 26 263 GRANZYME K.
FT ACT_SITE 66 66 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 115 115 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 213 213 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 51 67 BY SIMILARITY.
FT DISULFID 148 219 BY SIMILARITY.
FT DISULFID 180 198 BY SIMILARITY.
FT DISULFID 209 233 BY SIMILARITY.
SQ SEQUENCE 263 AA; 81951594F9AF08DE CRC64;

Query Match 74.0%; Score 37; DB 1; Length 263;
Best Local Similarity 71.4%; Pred. No. 5.1;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
DB 215 GPLICRG 221

RESULT 41

KLKR_PRANA STANDARD; PRT; 263 AA.
AC P32824;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE Glandular kallikrein, renal precursor (EC 3.4.21.35) (Tissue
DE kallikrein).
OS Praomys natalensis (African soft-furred rat) (Mastomys natalensis).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Rodentia; Sciurognathi; Muridae; Murinae;
OC Mastomys.
OX NCBI_TaxID=10112;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Salivary gland;
RX MEDLINE=94226702; PubMed=7909667;
RA Fahnstock M.;
RT "Characterization of kallikrein cDNAs from the African rodent
RT Mastomys";
RL DNA Cell Biol. 13:293-300(1994).
CC -! FUNCTION: GLANDULAR KALLIKREINS CLEAVE MET-LYS AND ARG-SER BONDS
CC IN KININOGEN TO RELEASE LYS-BRADYKININ.
CC -! CATALYTIC ACTIVITY: Preferential cleavage of Arg-I-Xaa bonds in
CC small molecule substrates. Highly selective action to release
CC kallidin (lysyl-bradykinin) from kininogen involves hydrolysis of
CC Met-I-Xaa or Leu-I-Xaa.
CC -! SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; X17352; CAA35232.1; -
DR PIR; S15686; S15686.
DR HSSP; P00757; 1SGF.
DR MEROPS; S01.160; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Multigene family; Zymogen;
KW SIGNAL 1 18 PROBABLE.
FT PROPEP 19 24 ACTIVATION PEPTIDE (PROBABLE).
FT CHAIN 25 263 GLANDULAR KALLIKREIN, RENAL.
FT ACT_SITE 65 65 CHARGE RELAY SYSTEM.
FT ACT_SITE 121 121 CHARGE RELAY SYSTEM.
FT ACT_SITE 215 215 CHARGE RELAY SYSTEM.
FT DISULFID 31 175 BY SIMILARITY.
FT DISULFID 50 66 BY SIMILARITY.
FT DISULFID 153 221 BY SIMILARITY.
FT DISULFID 186 200 BY SIMILARITY.
FT DISULFID 211 236 BY SIMILARITY.
FT CARBOHYD 102 102 N-LINKED (GLCNAC...) (PROBABLE).
SQ SEQUENCE 263 AA; 29130 MW; A8EB023B80037D5 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 263;
Best Local Similarity 66.7%; Pred. No. 5.1;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```

QY 1 GPLVCRGTL 9
Db 217 GPLICDGL 225

RESULT 42
GRAB_HUMAN STANDARD; PRT; 264 AA.
AC P49863;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Granzyme K precursor (EC 3.4.21.-) (Granzyme 3) (NK-tryptase-2)
DE (NK-TRYP-2) (Fragmentin 3).
DE GZMK OR TRYP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RC TISSUE=Asctites;
RX MEDLINE=95278340; PubMed=7758581;
RA Przetak M.M., Toast S., Schmidt B.F.;
RT "Cloning of cDNA for human granzyme 3.";
RL FEBS Lett. 364:268-271(1995).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Lymphocytes;
RX MEDLINE=96239310; PubMed=8656064;
RA Sayers T.J., Lloyd A.R., McVicar D.W., O'Connor M.D., Kelly J.M.,
RA Carter C.R.D., Wiltout T.A., Wiltout R.H., Smyth M.J.;
RT "Cloning and expression of a second human natural killer cell granule
RT tryptase, HNK-tryp-2/granzyme 3.";
RL J. Leukoc. Biol. 59:763-768(1996).
[3]
SEQUENCE OF 27-42, AND CHARACTERIZATION.
RX MEDLINE=89009866; PubMed=3262682;
RA Hameed A., Lowrey D.M., Lichtenheld M., Podack E.R.;
RT "Characterization of three serine esterases isolated from human IL-2
RT activated killer cells.";
RL J. Immunol. 141:3142-3147(1988).
[4]
PARTIAL SEQUENCE OF 27-48.
RX MEDLINE=93094755; PubMed=1460416;
RA Shi L., Kam C.-M., Powers J.C., Aebersold R., Greenberg A.H.;
RT "Purification of three cytotoxic lymphocyte granule serine proteases
RT that induce apoptosis through distinct substrate and target cell
RT interactions.";
RL J. Exp. Med. 176:1521-1529(1992).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC GRANULES.
CC -!- TISSUE SPECIFICITY: LUNG, SPLEEN, THYMUS, AND PERIPHERAL BLOOD
CC LEUKOCYTES.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPsin FAMILY. GRANZYME SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@lsb-sib.ch).
-----
DR EMBL; U35237; AAA79063.1;
DR EMBL; U26174; AAA4578.1;
DR HSSP; P20160; 1A7S.
DR MEROPS; S01.146; -.
DR MIM; 600784; -.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.

```

```

DR PROSITE; PS00134; TRYPSIN_HIS; 1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Zymogen; Signal.
FT SIGNAL 1 24 POTENTIAL.
FT PROPEP 25 26 ACTIVATION PEPTIDE.
FT CHAIN 27 264 GRANZYME K.
FT ACT_SITE 67 67 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 116 116 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 52 68 BY SIMILARITY.
FT DISULFID 149 220 BY SIMILARITY.
FT DISULFID 181 199 BY SIMILARITY.
FT DISULFID 210 234 BY SIMILARITY.
FT CONFLICT 34 34 S -> Q (IN REF. 4).
FT CONFLICT 43 43 S -> A (IN REF. 4).
SQ SEQUENCE 264 AA; 28882 MW; 2A93FDBAF9286CC5 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 264;
Best Local Similarity 71.4%; Pred. No. 5.1;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
Db 216 GPLICKG 222

RESULT 43
KLKD_HUMAN STANDARD; PRT; 277 AA.
ID KLKD_HUMAN Q9UKR3; Q9Y433;
AC Q9UKR3; Q9Y433;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Kallikrein 13 precursor (EC 3.4.21.-) (Kallikrein-like protein 4)
DE {KLK-L4}.
OS Homo sapiens (Human).
GN KLK13 OR KLK14.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
RX MEDLINE=20229789; PubMed=10766816;
RA Yousef G.M., Chang A., Diamandis E.P.;
RT "Identification and characterization of KLK-L4, a new kallikrein-like
RT gene that appears to be down-regulated in breast cancer tissues.";
RL J. Biol. Chem. 275:11891-11898(2000).
[2]
SEQUENCE FROM N.A.
RX MEDLINE=20229789; PubMed=10766816;
RA Lamerdin J.E., McGready P.M., Skowronski E., Viswanathan V.,
RA Burkhardt-Schultz K., Gordon L., Dias J., Ramirez M., Stilwagen S.,
RA Phan H., Velasco N., Do L., Regala W., Terry A., Brower A., Ganes J.,
RA Danganan L., Erler A., Christensen M., Georgescu A., Avila J., Liu S.,
RA Andreise T., Frankheim M., Attix C., Amico-Keller G., Coefield J.,
RA Duarte S., Lucas S., Bruce R., Thomas P., Quan G., Krommiller B.,
RA Arellano A., Sanders C., Ow D., Nolan M., Trong S., Kobayashi A.,
RA Olsen A.S., Carrano A.V.;
RT "Sequence analysis of chromosome 19ql3.4.";
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
[3]
SEQUENCE OF 1-180 FROM N.A.
RC TISSUE=Uterus;
RA Ansorge W., Wirkner U., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Secreted (Probable).
CC -!- TISSUE SPECIFICITY: EXPRESSED IN PROSTATE, BREAST, TESTIS AND
CC SALIVARY GLAND.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC TRYPSIN FAMILY. KALLIKREIN SUBFAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL; AF135024; AAD26425.2; -;
CC EMBL; AC011473; AAG23259.1; -;
CC EMBL; AL050220; CAB43320.1; ALT_INIT.
CC HSSP; P00763; 1DPO.
CC MTM; 605505; -;
CC InterPro; IPR001314; Chymotrypsin.
CC InterPro; IPR001254; Trypsin.
CC Pfam; PF00089; trypsin; 1.
CC PRINTS; PR00722; CHYMOTRYPSIN.
CC SMART; SM00020; TRYP_SPC; 1.
CC PROSITE; PS50240; TRYP_SIN_DOM; 1.
CC PROSITE; PS00134; TRYP_SIN_HIS; 1.
CC PROSITE; PS00135; TRYP_SIN_SER; 1.
KW Hydrolase; Serine protease; Glycoprotein; Signal.
FT SIGNAL 1 16 POTENTIAL.
FT CHAIN 17 277 KALLIKREIN 13.
FT ACT_SITE 76 76 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 124 124 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 218 218 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT DISULFID 42 178 BY SIMILARITY.
FT DISULFID 61 77 BY SIMILARITY.
FT DISULFID 157 224 BY SIMILARITY.
FT DISULFID 189 203 BY SIMILARITY.
FT DISULFID 214 239 BY SIMILARITY.
FT CARBOHYD 30 30 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 225 225 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 170 180 VNPPTLQCAN -> GMHPHRWPEAP (IN REF. 3).
SQ SEQUENCE 277 AA; 30570 MW; BABA9E8DCFB5D542 CRC64;

Query Match 74.0%; Score 37; DB 1; Length 277;
Best Local Similarity 77.8%; Pred. No. 5.3;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 GPLVCGRTL 9
| | | | |
DB 220 GPLVCGNRTL 228

RESULT 44
EL2B_HORSE
ID EL2B_HORSE STANDARD; PRT; 73 AA.
AC P27338;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Neutrophil elastase 2B (EC 3.4.21.-) (Proteinase 2B) (Fragments).
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE.
RC TISSUE=Neutrophils;
RX MEDLINE=94271153; PubMed=7516152;
RA Dublin A., Potempa J., Travis J.;
RT "Structural and functional characterization of elastases from horse neutrophils.";
RL Biochem. J. 300:401-406(1994).
CC -!- FUNCTION: MAY BE INVOLVED IN THE DEGRADATION OF CONNECTIVE TISSUE IN CHRONIC LUNG DISEASE.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE CC TRYPsin FAMILY. ELASTASE SUBFAMILY.
CC PIR; S44462; S44462.
DR HSSP; P20231; 1A00.
DR MEROPS; S01.131; -;
DR InterPro; IPR001254; Trypsin.
DR PROSITE; PS50240; TRYP_SIN_DOM; PARTIAL.
DR PROSITE; PS00134; TRYP_SIN_HIS; PARTIAL.

DR PROSITE; PS00135; TRYP_SIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON_CONS 31 32
FT NON_CONS 56 57
FT ACT_SITE 64 64
SQ SEQUENCE 73 AA; 7615 MW; AFFB0B330DB69041 CRC64;
CHARGE RELAY SYSTEM.

Query Match 72.0%; Score 36; DB 1; Length 73;
Best Local Similarity 85.7%; Pred. No. 2.3;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 1 GPLVCG 7
| | | | |
DB 66 GPLVCG 72

RESULT 45
DDN1_BOVIN
ID DDN1_BOVIN STANDARD; PRT; 226 AA.
AC P80219;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update).
DE Duodenase I (EC 3.4.21.-) (Duodenum serine protease).
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE.
RC TISSUE=Duodenum;
RX MEDLINE=95172076; PubMed=7867649;
RA Zamolodchikova T.I., Vorotyntseva T.I., Nazimov I.V., Grishina G.A.;
RT "Duodenase, a new serine protease of unusual specificity from bovine duodenal mucosa. Primary structure of the enzyme.";
RL Eur. J. Biochem. 227:873-879(1995).
RN [2]
RP SEQUENCE OF 1-20 AND 172-183.
RX MEDLINE=93048618; PubMed=1425193;
RA Antonov V.K., Vorotyntseva T.I., Zamolodchikova T.S.;
RT "Duodenase -- a new serine proteinase with unusual specificity.";
RL Dokl. Akad. Nauk SSSR 324:1318-1322(1992).
RN [3]
RP SEQUENCE OF 1-24.
RX MEDLINE=95172075; PubMed=7867648;
RA Zamolodchikova T.S., Vorotyntseva T.I., Antonov V.K.;
RT "Duodenase, a new serine protease of unusual specificity from bovine duodenal mucosa. Purification and properties.";
RL Eur. J. Biochem. 227:866-872(1995).
RN [4]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
RX MEDLINE=9402518; PubMed=10944388;
RA Pletnev V.Z., Zamolodchikova T.S., Pangborn W.A., Duax W.L.;
RT "Crystal structure of bovine duodenase, a serine protease, with dual trypsin and chymotrypsin-like specificities.";
RL Proteins 41:8-16(2000).
CC -!- FUNCTION: Protease which has both trypsin-like and chymotrypsin-like activities.
CC -!- CATALYTIC ACTIVITY: PREFERENTIAL CLEAVAGE: LYS-, ARG-, TYR-, PHE-, LEU-
CC -!- SUBUNIT: MONOMER.
CC -!- MISCELLANEOUS: THE OPTIMUM PH AND TEMPERATURE OF DUODENASE I ARE 8.0 AND 50 DEGREES CELSIUS, RESPECTIVELY.
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE CC TRYPsin FAMILY.
CC HSSP; P04187; 2CPI.
DR MEROPS; S01.142; -;
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; TRYP_SPC; 1.

DR PROSITE; PS0240; TRYPSIN_DOM; 1.
 DR PROSITE; PS00134; TRYPSIN_HIS; 1.
 DR PROSITE; PS00135; TRYPSIN_SER; 1.
 KW Hydrolase; Serine protease; Glycoprotein.
 FT ACT_SITE 44 44 CHARGE RELAY SYSTEM.
 FT ACT_SITE 87 87 CHARGE RELAY SYSTEM.
 FT ACT_SITE 181 181 CHARGE RELAY SYSTEM.
 FT DISULFID 29 45
 FT DISULFID 121 187
 FT DISULFID 152 166
 FT CARBOHYD 50 50 N-LINKED (GLCNAC...).
 SQ SEQUENCE 226 AA; 25051 MW; A4CA2CE66736CD12 CRC64;

Query Match 72.0%; Score 36; DB 1; Length 226;
 Best Local Similarity 85.7%; Pred. No. 6.8;
 Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
 |||||
 Db 183 GPLVCNG 189

Search completed: November 6, 2002, 12:09:16
 Job time : 6.33333 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 11 seconds
(without alignments)
78.619 Million cell updates/sec

Title: US-09-905-083-80

Perfect score: 50

Sequence: 1 GPLVCRGTL 9

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR-71.*

1: piri.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	50	100.0	253	2 A53968	serine proteinase
2	40	80.0	261	1 A32297	semenogelase (EC 3
3	40	80.0	261	2 A29586	tissue kallikrein
4	40	80.0	262	2 T35999	probable aminoglyc
5	39	78.0	155	2 G72548	hypothetical prote
6	39	78.0	265	1 KQTP	tissue kallikrein
7	38	76.0	246	1 DBHU	complement factor
8	38	76.0	261	1 S35711	semenogelase (EC 3
9	37	74.0	46	2 I49416	glandular kallikre
10	37	74.0	96	2 A05308	tissue kallikrein
11	37	74.0	104	2 S15395	tissue kallikrein-
12	37	74.0	156	2 B23863	tissue kallikrein
13	37	74.0	188	2 B32340	tissue kallikrein
14	37	74.0	225	2 S45356	probable serine pr
15	37	74.0	239	2 A27207	tissue kallikrein
16	37	74.0	240	2 S39047	trypsin (EC 3.4.21
17	37	74.0	241	2 S39048	trypsin (EC 3.4.21
18	37	74.0	244	2 A44284	tissue kallikrein
19	37	74.0	250	2 S15685	kallikrein, glandu
20	37	74.0	254	2 S35585	chymotrypsin-like
21	37	74.0	258	2 I56220	trypsin (EC 3.4.21-
22	37	74.0	259	1 KQRTTN	tissue kallikrein
23	37	74.0	259	2 D23863	tissue kallikrein
24	37	74.0	259	2 A29746	neutrophin - mouse
25	37	74.0	260	2 I56559	tissue kallikrein
26	37	74.0	261	1 KQMSI	tissue kallikrein
27	37	74.0	261	1 NGMSG	7S nerve growth fa
28	37	74.0	261	1 TRMSM5	tissue kallikrein
29	37	74.0	261	2 A34079	tissue kallikrein

30	37	74.0	261	2 S01971	tissue kallikrein
31	37	74.0	261	2 A28062	gamma-renin (EC 3.
32	37	74.0	261	2 A25606	tissue kallikrein
33	37	74.0	261	2 A29745	tissue kallikrein
34	37	74.0	261	2 S45303	tissue kallikrein
35	37	74.0	263	2 S15686	tissue kallikrein
36	37	74.0	264	2 S65663	granzyme 3 (EC 3.4
37	37	74.0	599	2 T12994	riboflavin biosynt
38	36	72.0	66	2 I52972	kallikrein - mouse
39	36	72.0	73	2 S44462	elastase (EC 3.4.2
40	36	72.0	149	1 KQMSM	tissue kallikrein
41	36	72.0	226	2 S69370	duodenase - bovine
42	36	72.0	229	1 TRDFS	trypsin (EC 3.4.21
43	36	72.0	231	1 TRPGTR	trypsin (EC 3.4.21
44	36	72.0	231	2 S31778	trypsin (EC 3.4.21
45	36	72.0	236	1 A32121	snake venom factor

ALIGNMENTS

RESULT 1

A53968

serine proteinase SCCE precursor - human

N;Alternate names: stratum corneum chymotryptic enzyme

C;Species: Homo sapiens (man)

C;Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999

C;Accession: A53968

R;Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstei, A.; Egelrud

J. Biol. Chem. 269, 19420-19426, 1994

A;Title: Cloning, expression, and characterization of stratum corneum chymotryptic en

A;Reference number: A53968; MUID:94308225

A;Accession: A53968

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-253 <HAN>

A;Cross-references: GB:L33404; NID:g521214; PIDN:AAC37551.1; PID:g532504

C;Genetics:

A;Gene: GDB:PRSS6; SCCE

A;Cross-references: GDB:377730

A;Map position: 7q35-7q35

C;Superfamily: trypsin; trypsin homology

F;30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 50; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.086;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

| | | | | | | | | |

Db 207 GPLVCRGTL 215

RESULT 2

A32297

semenogelase (EC 3.4.21.77) precursor [validated] - human

N;Alternate names: gamma-seminoprotein; p-30 antigen; prostate-specific antigen; pro.

C;Species: Homo sapiens (man)

C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 08-Dec-2000

C;Accession: A32297; A32423; S03604; S05468; S05467; A32546; S02239; A26757; C31567;

R;Riegman, P.H.J.; Vlietstra, R.J.; van der Korput, J.A.G.M.; Romijn, J.C.; Trapman,

Biochem. Biophys. Res. Commun. 159, 95-102, 1989

A;Title: Characterization of the prostate-specific antigen gene: a novel human kallik

A;Reference number: A32297; MUID:89165891

A;Accession: A32297

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-261 <RIE>

A;Cross-references: GB:M24543

R;Lundwall, A.

Biochem. Biophys. Res. Commun. 161, 1151-1159, 1989

A;Title: Characterization of the gene for prostate-specific antigen, a human glandul

A;Reference number: A32423; MUID:89302090

A:Accession: A32423
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-261 <LUN>
A:Cross-references: GB:M27274; NID:g190552; PIDN:AAA60192.1; PID:g190553
R:Digby, M.; Zhang, X.Y.; Richards, R.I.
Nucleic Acids Res. 17, 2137, 1989
A:Title: Human prostate specific antigen (PSA) gene: structure and linkage to the kallikrein gene
A:Reference number: S03604; MUID:89183632
A:Accession: S03604
A:Molecule type: DNA
A:Residues: 1-261 <DIG>
A:Cross-references: EMBL:X13940
R:Klobeck, H.G.; Combratio, G.; Schulz, P.; Arbusow, V.; Fittler, F.
submitted to the EMBL Data Library, May 1989
A:Reference number: S05468
A:Accession: S05468
A:Molecule type: DNA
A:Residues: 1-261 <KL1>
A:Cross-references: EMBL:X14810; NID:g35732; PIDN:CAA32915.1; PID:g296671
R:Klobeck, H.G.; Combratio, G.; Schulz, P.; Arbusow, V.; Fittler, F.
Nucleic Acids Res. 17, 3981, 1989
A:Title: Genomic sequence of human prostate specific antigen (PSA).
A:Reference number: S05467; MUID:89282407
A:Accession: S05467
A:Molecule type: DNA
A:Residues: 1-29 <KL2>
A:Cross-references: EMBL:X14810
R:Henttu, P.; Viikko, P.
Biochem. Biophys. Res. Commun. 160, 903-910, 1989
A:Title: cDNA coding for the entire human prostate specific antigen shows high homology to the kallikrein gene
A:Reference number: A32546; MUID:89246551
A:Accession: A32546
A:Molecule type: mRNA
A:Residues: 1-72, 'T', 74-85, 'I', 87-174, 'P', 176-183, 'Q', 185-259, 'D', 261 <HEN>
A:Cross-references: GB:M26663
R:Schulz, P.; Stucke, R.; Feldmann, H.; Combratio, G.; Klobeck, H.G.; Fittler, F.
Nucleic Acids Res. 16, 6226, 1988
A:Title: Sequence of a cDNA clone encompassing the complete mature human prostate specific antigen
A:Reference number: S02239; MUID:88289366
A:Accession: S02239
A:Molecule type: mRNA
A:Residues: 17-63, 'T', 65-135, 'M', 137-261 <SCH>
A:Cross-references: EMBL:X07730
R:Lundwall, A.; Lilja, H.
FEBS Lett. 214, 317-322, 1987
A:Title: Molecular cloning of human prostate specific antigen cDNA.
A:Reference number: A26757; MUID:87190978
A:Accession: A26757
A:Molecule type: mRNA
A:Residues: 5-261 <LU2>
R:Riegman, P.H.J.; Klaassen, P.; van der Korp, J.A.G.M.; Romijn, J.C.; Trapman, J.
Biochem. Biophys. Res. Commun. 155, 181-188, 1988
A:Title: Molecular cloning and characterization of novel prostate antigen cDNA's.
A:Reference number: A90144; MUID:88326297
A:Accession: C31567
A:Molecule type: mRNA
A:Residues: 5-261 <RI2>
A:Cross-references: GB:M21895; NID:g189523; PIDN:AAA59995.1; PID:g189524
A:Note: the authors translated the codon GGC for residue 28 as Arg and TGG for residue 29 as Arg
A:Accession: A31567
A:Molecule type: mRNA
A:Residues: 5-210, 'WVILITELTPA', 223, 'PMLHGLSLV', 233, 'WRGVV' <RI3>
A:Cross-references: GB:M21897; NID:g189529; PIDN:AAA59996.1; PID:g189526
A:Note: the authors translated the codon GGC for residue 28 as Arg and TGG for residue 29 as Arg
A:Accession: B31567
A:Molecule type: mRNA
A:Residues: 5-164, 'CTPGDGAAGSPDAWV' <RI4>
A:Cross-references: GB:M21897; NID:g189529; PIDN:AAA59997.1; PID:g189530
A:Note: the authors translated the codon GGC for residue 28 as Arg, TGC for residue 29 as Arg
R:Schaller, J.; Akiyama, K.; Tsuda, R.; Hara, M.; Marti, T.; Rickli, E.E.
Eur. J. Biochem. 170, 111-120, 1987
A:Title: Isolation, characterization and amino-acid sequence of gamma-seminoprotein, a

A:Reference number: S00232; MUID:88082806
A:Accession: S00232
A:Molecule type: protein
A:Residues: 25-261 <SC2>
R:Watt, K.W.K.; Lee, P.J.; M'Imkulu, T.; Chan, W.P.; Loo, R.
Proc. Natl. Acad. Sci. U.S.A. 83, 3166-3170, 1986
A:Title: Human prostate-specific antigen: structural and functional similarity with s
A:Reference number: A23937; MUID:86205857
A:Accession: A23937
A:Molecule type: protein
A:Residues: 25-93, 'I', 95-164, 'HL', 166, 'YDQM', 169-174, 'Q', 176-261 <WAT>
R:Moreno, J.M.
submitted to the EMBL Data Library, November 1994
A:Reference number: G07735
A:Accession: G01551
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-261 <MOR>
A:Cross-references: EMBL:U17040; NID:g595945; PIDN:AAA56764.1; PID:g595946
R:Monne, M.; Croce, C.M.; Yu, H.; Diamandis, E.P.
Cancer Res. 54, 6344-6347, 1994
A:Title: Molecular characterization of prostate-specific antigen messenger RNA expres
A:Reference number: I52712; MUID:95079406
A:Accession: I52712
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 'POAVLHPESCVT', 1-176 <RES>
A:Cross-references: GB:S75755; NID:g861469; PIDN:AAI41485.1; PID:g4261885
R:Christensson, A.; Lilja, H.
Eur. J. Biochem. 220, 45-53, 1994
A:Title: Complex formation between protein C inhibitor and prostate-specific antigen
A:Reference number: S41212; MUID:94164172
A:Accession: S41212
A:Molecule type: protein
A:Residues: 25-30, 'X', 32-49 <CHR>
C:Comment: This enzyme preferentially cleaves after tyrosine residues.
C:Genetics:
A:Gene: GDB:APS; PSA
A:Cross-references: GDB:119695; OMIM:176820
A:Map position: 19q13.3-19q13.3
A:Introns: 16/1; 69/2; 165/1; 210/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; prostate; serine proteinase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-24/Domain: propeptide #status predicted <PRO>
F:25-261/Product: semenogelase #status experimental <MAT>
F:25-253/Domain: trypsin homology <TRY>
F:65,120,213/Active site: His, Asp, Ser #status predicted

Query Match 80.0%; Score 40; DB 1; Length 261;
Best Local Similarity 77.8%; Pred. No. 6.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
DB 215 GPLVCNGVL 223
|||||
I I I I I I I

RESULT 3
A29586
tissue kallikrein (EC 3.4.21.35) hGK-1 precursor - human
N:Alternate names: glandular kallikrein
C:Species: Homo sapiens (man)
C:Date: 08-Mar-1989 #sequence_revision 08-Mar-1989 #text_change 22-Jun-1999
C:Accession: A29586
R:Schedlich, L.J.; Bennetts, B.H.; Morris, B.J.
DNA 6, 429-437, 1987
A:Title: Primary structure of a human glandular kallikrein gene.
A:Reference number: A29586; MUID:88054467
A:Accession: A29586
A:Molecule type: DNA
A:Residues: 1-261 <SCH>
A:Cross-references: GB:M18157; NID:g186640; PIDN:AAA74454.1; PID:g386842

A:Note: the authors translated the codon TAC for residue 43 as Trp

```
C:Genetics:
A:Introns: 16/1; 69/2; 165/1; 210/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:25-253/Domain: trypsin homology <TRY>
F:65,120,213/Active site: His, Asp, Ser #status predicted

Query Match      80.0%  Score 40;  DB 2;  Length 261;
Best Local Similarity  77.8%  Pred. No. 6.1;
Matches  7;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

QY  1  GPLVCRGTL  9
    ||||| 1
Db   215  GPLVCGNVL 223

RESULT 4
T35999
C:Species: Streptomyces coelicolor
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 26-May-2000
C:Accession: T35999
R:Seeger, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, July 1999
A:Reference number: Z21574
A:Accession: T35999
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-262 <SEE>
A:Cross-references: EMBL:AL096839; PIDN:CAB50752.1; GSPDB:GN00070; SCOPDB:SCC22.09
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCOPDB:SCC22.09
C:Superfamily: Escherichia coli aminoglycoside N3'-acetyltransferase

Query Match      80.0%  Score 40;  DB 2;  Length 262;
Best Local Similarity  88.9%  Pred. No. 6.1;
Matches  8;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

QY  1  GPLVCRGTL  9
    ||||| 1
Db   8  GPLVTRGTL 16

RESULT 5
G72548
hypothetical protein APE1675 - Aeropyrum pernix (strain K1)
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999
C:Accession: G72548
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takahawa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kana, R.; 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyrum pernix
A:Reference number: A72450; MUID:99310339
A:Accession: G72548
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-155 <RAW>
A:Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80676.1; PID:d1044462; PID:g5105244
A:Experimental source: strain K1
C:Genetics:
A:Gene: APE1675

Query Match      78.0%  Score 39;  DB 2;  Length 155;
Best Local Similarity  77.8%  Pred. No. 5.8;
Matches  7;  Conservative  1;  Mismatches  1;  Indels  0;  Gaps  0;

QY  1  GPLVCRGTL  9
    ||| 1
Db   23  GPLITRGTL 31
```

RESULT 6

KORTP

tissue kallikrein (EC 3.4.21.35) precursor - rat
N:Alternate names: glandular kallikrein; kininogenin; true tissue kallikrein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 05-Apr-1983 #sequence_revision 05-Apr-1983 #text_change 16-Jun-2000
C:Accession: A00944; A1429; A25137; JX0073; A23863; A33359
R:Swift, G.H.; Dagorn, J.C.; Ashley, P.L.; Cummings, S.W.; MacDonald, R.J.
Proc. Natl. Acad. Sci. U.S.A. 79, 7263-7267, 1982
A:Title: Rat pancreatic kallikrein mRNA: nucleotide sequence and amino acid sequence
A:Reference number: A00944; MUID:83117659
A:Accession: A00944
A:Molecule type: mRNA
A:Residues: 1-265 <SWI>
A:Experimental source: pancreatic
R:Kato, H.; Nakanishi, E.; Enjyoji, K.; Hayashi, I.; Oh-Ishi, S.; Iwanaga, S.
J. Biochem. 102, 1389-1404, 1987
A:Title: Characterization of serine proteinases isolated from rat submaxillary gland
A:Reference number: A1429; MUID:88198057
A:Accession: A1429
A:Status: preliminary
A:Molecule type: protein
A:Residues: 29-53; 'X', 55-87 <KAT>
R:Gerald, W.L.; Chao, J.; Chao, L.
Biochim. Biophys. Acta 866, 1-14, 1986
A:Title: Immunological identification of rat tissue kallikrein cDNA and characterization
A:Reference number: A25137; MUID:86131678
A:Accession: A25137
A:Molecule type: mRNA
A:Residues: 115-265 <GER>
R:Inoue, H.; Fukui, K.; Miyake, Y.
J. Biochem. 105, 834-840, 1989
A:Title: Identification and structure of the rat true tissue kallikrein gene expressed
A:Reference number: JX0073; MUID:89327211
A:Accession: JX0073
A:Molecule type: DNA
A:Residues: 1-265 <INO>
A:Cross-references: GB:D00448; NID:g220792; PIDN:BAA00346.1; PID:g220794
A:Experimental source: kidney
R:Ashley, P.L.; MacDonald, R.J.
Biochemistry 24, 4512-4520, 1985
A:Title: Kallikrein-related mRNAs of the rat submaxillary gland: nucleotide sequence
A:Reference number: A23863; MUID:86051477
A:Accession: A23863
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-265 <ASH>
A:Cross-references: GB:M11563; NID:g205029; PIDN:AAA41464.1; PID:g205030
A:Experimental source: submaxillary gland
R:Wines, D.R.; Brady, J.M.; Pritchett, D.B.; Roberts, J.L.; MacDonald, R.J.
J. Biol. Chem. 264, 7653-7662, 1989
A:Title: Organization and expression of the rat kallikrein gene family.
A:Reference number: A33359; MUID:89214217
A:Accession: A33359
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 5-265 <WIN>
A:Cross-references: GB:M23874; GB:J04701; GB:M23875; GB:M23876; NID:g205007; PIDN:AA.
C:Comment: The kallikreins liberate lysyl-bradykinin, a vasoactive decapeptide, from
C:Comment: The protein presumably assumes the two-chain form by cleavage between res
C:Comment: Tissue kallikreins cleave Met-Lys and Arg-Ser bonds in kininogen to relea
C:Genetics:
A:Introns: 20/1; 73/2; 169/1; 214/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; serine proteinase; zymogen
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-28/Domain: activation peptide #status predicted <APT>
F:29-265/Product: tissue kallikrein, pancreatic #status predicted <MPT>
F:29-257/Domain: trypsin homology <TRY>
F:35-177,54-70,156-223,188-202,213-238/Disulfide bonds: #status predicted
F:69,124,217/Active site: His, Asp, Ser #status predicted

Query Match 78.0%; Score 39; DB 1; Length 265;
 Best Local Similarity 66.7%; Pred. No. 9.5;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 |||||

Db 219 GPLVCGVGL 227

RESULT 7
 DBHU
 complement factor D (EC 3.4.21.46) precursor [validated] - human (fragment)
 N;Alternate names: adipsin; C3 convertase activator
 C;Species: Homo sapiens (man)
 C;Date: 28-Aug-1985 #sequence_revision 31-Dec-1992 #text_change 08-Dec-2000
 C;Accession: A40197; A00936; A60571; S66645
 R;White, R.T.; Damm, D.; Hancock, N.; Rosen, B.S.; Lowell, B.B.; Usher, P.; Flier, J.S.;
 J. Biol. Chem. 267, 9210-9213, 1992
 A;Title: Human adipsin is identical to complement factor D and is expressed at high level
 A;Reference number: A40197; MUID:92250520
 A;Accession: A40197
 A;Molecule type: mRNA
 A;Residues: 1-246 <WHI>
 A;CROSS-references: GB:M84526
 R;Niemann, M.A.; Shown, A.S.; Bennett, J.C.; Volanakis, J.E.
 Biochemistry 23, 2482-2486, 1984
 A;Title: Amino acid sequence of human D of the alternative complement pathway.
 A;Reference number: A00936; MUID:85000441
 A;Accession: A00936
 A;Molecule type: protein
 A;Residues: 19-44,'G',46-51,'Q',53-75,'TH',78,'P',80-83,'XXXITIE',90-172,86-91,185-235,'
 A;Note: a few residues were assigned from the previously published sequence of Reid et al
 R;Miyata, T.; Oda, O.; Inagi, R.; Sugiyama, S.; Miyama, A.; Maeda, K.; Nakashima, I.; Ya
 Mol. Immunol. 27, 637-644, 1990
 A;Title: Molecular and functional identification and purification of complement component
 A;Reference number: A60571; MUID:90370044
 A;Accession: A60571
 A;Molecule type: protein
 A;Residues: 19-20,'XX',23-27,'XX',30-31,'XX',34,'X',36-40 <MIY>
 R;Balke, N.; Holtkamp, U.; Hoerl, W.H.; Tschesche, H.
 FEBS Lett. 371, 300-302, 1995
 A;Title: Inhibition of degranulation of human polymorphonuclear leukocytes by complement
 A;Reference number: S66645; MUID:96013156
 A;Accession: S66645
 A;Status: preliminary
 A;Molecule type: protein
 A;Residues: 19-44,'C',46-48 <BAL>
 C;Comment: Factor D cleaves factor B when the latter is complexed with factor C3b, activ
 C;Genetics:
 A;Gene: GDB:DF
 A;CROSS-references: GDB:I32645; OMIM:134350
 A;Map position: Xpter-Xqter
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: complement alternate pathway; hydrolase; plasma; serine proteinase
 F;1-18/Domain: signal sequence #status predicted <SIG>
 F;19-246/Product: complement factor D (fragment) #status experimental <MAT>
 F;19-241/Domain: complement factor D (fragment) #status predicted
 F;44-60,141-207,172-188,197-222/Disulfide bonds: #status predicted
 F;59,105,201/Active site: His, Asp, Ser #status predicted

Query Match 76.0%; Score 38; DB 1; Length 246;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 |||||

Db 203 GPLVCGVGL 211

RESULT 8
 S35711
 semenogelase (EC 3.4.21.77) precursor - rhesus macaque
 N;Alternate names: gamma-seminoprotein; P-30 antigen; prostate-specific antigen; prostat

C;Species: Macaca mulatta (rhesus macaque)
 C;Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 19-May-2000
 C;Accession: S35711; S34239
 R;Gauthier, E.R.; Chapdelaine, P.; Tremblay, R.R.; Dube, J.Y.
 Biochim. Biophys. Acta 1174, 207-210, 1993
 A;Title: Characterization of rhesus monkey prostate specific antigen cDNA.
 A;Reference number: S35711; MUID:93363642
 A;Accession: S35711
 A;Molecule type: mRNA
 A;Residues: 1-261 <GAU>
 A;CROSS-references: EMBL:X73560; NID:9311843; PIDN:CAA51957.1; PID:9311844
 C;Comment: This enzyme preferentially cleaves after tyrosine residues.
 C;Superfamily: trypsin; trypsin homology
 C;Keywords: glycoprotein; hydrolase; prostate; serine proteinase
 F;1-17/Domain: signal sequence #status predicted <SIG>
 F;18-24/Domain: propeptide #status predicted <PRO>
 F;25-261/Product: semenogelase #status predicted <MAT>
 F;25-253/Domain: trypsin homology <TRY>
 F;65,120,213/Active site: His, Asp, Ser #status predicted

Query Match 76.0%; Score 38; DB 1; Length 261;
 Best Local Similarity 77.8%; Pred. No. 14;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 |||||

Db 215 GPLVCGVGL 223

RESULT 9
 I49416
 glandular kallikrein - western wild mouse (fragment)
 C;Species: Mus spretus (western wild mouse)
 C;Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 22-Jun-1999
 C;Accession: I49416
 R;Ko, M.S.; Wang, X.; Horton, J.H.; Hagen, M.D.; Takahashi, N.; Maezaki, Y.; Nadeau,
 Mamm. Genome 5, 349-355, 1994
 A;Title: Genetic mapping of 40 cDNA clones on the mouse genome by PCR.
 A;Reference number: I48934; MUID:94319082
 A;Accession: I49416
 A;Status: preliminary; translated from GB/EMBL/DBJ
 A;Molecule type: DNA
 A;Residues: 1-46 <RES>
 A;CROSS-references: EMBL:U05716; NID:g497047; PIDN:AAB60472.1; PID:g497048
 C;Superfamily: trypsin; trypsin homology

Query Match 74.0%; Score 37; DB 2; Length 46;
 Best Local Similarity 66.7%; Pred. No. 4.5;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 |||||

Db 1 GPLICDGLV 9

RESULT 10
 A05308
 tissue kallikrein (EC 3.4.21.35), submandibular mCK-2 - mouse (fragment)
 N;Alternate names: glandular kallikrein
 C;Species: Mus musculus (house mouse)
 C;Date: 05-Jun-1987 #sequence_revision 05-Jun-1987 #text_change 22-Jun-1999
 C;Accession: A05308
 R;Nason, A.J.; Evans, B.A.; Cox, D.R.; Shine, J.; Richards, R.I.
 Nature 303, 300-307, 1983
 A;Title: Structure of mouse kallikrein gene family suggests a role in specific proces
 A;Reference number: A00941; MUID:83219214
 A;Accession: A05308
 A;Molecule type: DNA
 A;Residues: 1-96 <MAS>
 A;CROSS-references: GB:V00829; NID:952775; PIDN:CAA24212.1; PID:g388523
 A;Experimental source: Quakenbush inbred strain
 A;Note: This sequence has been translated from two exons (11-147 and 522-674) located
 C;Comment: Glandular kallikreins cleave Met-Lys and Arg-Ser bonds in kininogen to rel

A;Residues: 1-239 <DUN>
A;Note: 50-Trip was also found

C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; serine proteinase
F;1-231/Domain: trypsin homology <TRY>
F;41,96,191/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 239;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 193 GPLICDGV 201

RESULT 16

S39047

trypsin (EC 3.4.21.4) I - Atlantic cod

C;Species: Gadus morhua (Atlantic cod)

C;Date: 20-May-1994 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000

C;Accession: S39047

R;Gudmundsdottir, A.; Gudmundsdottir, E.; Oskarsson, S.; Bjarnason, J.B.; Eakin, A.K.; C

Eur. J. Biochem. 217, 1091-1097, 1993

A;Title: Isolation and characterization of cDNAs from Atlantic cod encoding two different

A;Reference number: S39047; MUID:94039130

A;Accession: S39047

A;Molecule type: mRNA

A;Residues: 1-240 <GUD>

A;Cross-references: EMBL:X76886; NID:g450517; PIDN:CAA54214.1; PID:g1334752

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; serine proteinase

F;19-233/Domain: trypsin homology <TRY>

F;58,102,194/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 240;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 196 GPVVCNGVL 204

RESULT 17

S39048

trypsin (EC 3.4.21.4) X - Atlantic cod

C;Species: Gadus morhua (Atlantic cod)

C;Date: 20-May-1994 #sequence_revision 03-Aug-1995 #text_change 20-Jun-2000

C;Accession: S39048

R;Gudmundsdottir, A.; Gudmundsdottir, E.; Oskarsson, S.; Bjarnason, J.B.; Eakin, A.K.; C

Eur. J. Biochem. 217, 1091-1097, 1993

A;Title: Isolation and characterization of cDNAs from Atlantic cod encoding two different

A;Reference number: S39047; MUID:94039130

A;Accession: S39048

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-241 <GUD>

A;Cross-references: EMBL:X76887; NID:g450519; PIDN:CAA54215.1; PID:g1334753

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; serine proteinase

F;20-234/Domain: trypsin homology <TRY>

F;59,103,195/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 241;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 197 GPVVCNGVL 205

RESULT 18

A44284

tissue kallikrein (EC 3.4.21.35) homolog - rat (fragment)

N;Alternate names: proteinase B light chain

C;Species: Rattus norvegicus (Norway rat)

C;Date: 19-Nov-1993 #sequence_revision 18-Nov-1994 #text_change 22-Jun-1999

C;Accession: A44284; C41429

R;Ma, J.X.; Chao, J.; Chao, L.

Biochemistry 31, 10922-10928, 1992

A;Title: Molecular cloning and characterization of rKlk10, a cDNA encoding T-kininoge

A;Reference number: A44284; MUID:93041794

A;Accession: A44284

A;Status: preliminary

A;Molecule type: nucleic acid

A;Residues: 1-244 <MAL>

A;Cross-references: GB:S48142; NID:g259430; PIDN:AAB24071.1; PID:g259431

A;Experimental source: submandibular gland

A;Note: Sequence extracted from NCBI backbone (NCBIN:118095, NCBIPI:118096)

R;Kato, H.; Nakanishi, E.; Enyoji, K.; Hayashi, I.; Oh-Ishi, S.; Iwanaga, S.

J. Biochem. 102, 1389-1404, 1987

A;Title: Characterization of serine proteinases isolated from rat submaxillary gland;

A;Reference number: A41429; MUID:88198057

A;Accession: C41429

A;Status: preliminary

A;Molecule type: protein

A;Residues: 10-61 <KAT>

C;Superfamily: trypsin; trypsin homology

C;Keywords: hydrolase; serine proteinase

F;10-236/Domain: trypsin homology <TRY>

F;48,103,196/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 244;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 198 GPLICDGV 206

RESULT 19

S15685

kallikrein, glandular - multimammate rat (Mastomys natalensis) (fragment)

C;Species: Mastomys natalensis

C;Date: 09-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 20-Jun-2000

C;Accession: S15685

R;Bell, R.A.; Fahnestock, M.

submitted to the EMBL Data Library, December 1989

A;Description: Sequence of a cDNA clone coding for a glandular kallikrein from Mastom

A;Reference number: S15685

A;Accession: S15685

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-250 <BEL>

A;Cross-references: EMBL:X17351; NID:g555524; PIDN:CAA35231.1; PID:g1334162

C;Superfamily: trypsin; trypsin homology

F;22-250/Domain: trypsin homology <TRY>

F;62,117,211/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 250;
Best Local Similarity 66.7%; Pred. No. 21;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 213 GPLICDGV 221

RESULT 20

S35585

chymotrypsin-like proteinase (EC 3.4.21.-) precursor - California red abalone

C;Species: Haliotis rufescens (California red abalone)

C;Date: 07-Apr-1994 #sequence_revision 19-May-1994 #text_change 22-Jun-1999

C;Accession: S35585; S32750
R;Groppe, J.C.; Morse, D.E.
Arch. Biochem. Biophys. 305, 159-169, 1993
A;Title: Molluscan chymotrypsin-like protease: structure, localization, and substrate sp
A;Reference number: S35585; MUID:93343624
A;Accession: S35585
A;Molecule type: mRNA
A;Residues: 1-254 <GRO>
A;Cross-references: EMBL:X71438; NID:g296361; PIDN:CAA50572.1; PID:g296362
C;Superfamily: trypsin; trypsin homology
C;Keywords: hydrolase; intestine; serine proteinase; zymogen
F;1-23/Domain: signal sequence #status predicted <SIG>
F;23-254/Product: chymotrypsin-like proteinase #status predicted <MAT>
F;24-251/Domain: trypsin homology <TRY>

Query Match 74.0%; Score 37; DB 2; Length 254;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||||| I
Db 214 GPLVCGNTL 222

RESULT 21
I56220
trypsin 2 - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 22-Jun-1999
C;Accession: I56220
R;Sayers, T.J.; Wiltrout, T.A.; Smyth, M.J.; Ottaway, K.S.; Pilaro, A.M.; Sowder, R.; He
J. Immunol. 152, 2289-2297, 1994
A;Title: Purification and cloning of a novel serine protease, RNK-tryp-2, from the gran
A;Reference number: I56220; MUID:94179809
A;Accession: I56220
A;Status: preliminary; translated from GB/EMBL/DDBU
A;Molecule type: mRNA
A;Residues: 1-258 <RES>
A;Cross-references: GB:I19694; NID:g487878; PIDN:AAA42057.1; PID:g487879
C;Superfamily: trypsin; trypsin homology
F;26-248/Domain: trypsin homology <TRY>

Query Match 74.0%; Score 37; DB 2; Length 258;
Best Local Similarity 71.4%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
||||| I
Db 210 GPLVCRG 216

RESULT 22
KQRTN
tonin (EC 3.4.21.-) precursor - rat
N;Alternate names: esterase 1
C;Species: Rattus norvegicus (Norway rat)
C;Date: 25-Feb-1985 #sequence_revision 02-Dec-1994 #text_change 18-Jun-1999
C;Accession: B33359; A32340; C23863; A93223; A94635; A34050; A30971; A00945
R;Wines, D.R.; Brady, J.M.; Pritchett, D.B.; Roberts, J.L.; Macdonald, R.J.
J. Biol. Chem. 264, 7653-7662, 1989
A;Title: Organization and expression of the rat kallikrein gene family.
A;Reference number: A33359; MUID:89214217
A;Accession: B33359
A;Molecule type: DNA
A;Residues: 1-259 <WIN>
A;Cross-references: GB:M23877; GB:J04702; GB:M23878; NID:g207411; PIDN:AAA42259.1; PID:g
R;Shai, S.Y.; Woodley-Miller, C.; Chao, J.; Chao, L.
Biochemistry 28, 5334-5343, 1989
A;Title: Characterization of genes encoding rat tonin and a kallikrein-like serine prote
A;Reference number: A32340; MUID:89375248
A;Accession: A32340
A;Molecule type: DNA
A;Residues: 1-259 <SHA>

A;Cross-references: GB:M26533; NID:g206775; PIDN:AAA42081.1; PID:g206776; GB:J02860
R;Ashley, P.L.; Macdonald, R.J.
Biochemistry 24, 4512-4520, 1985
A;Title: Kallikrein-related mRNAs of the rat submaxillary gland: nucleotide sequence
A;Reference number: A23863; MUID:86051477
A;Accession: C23863
A;Molecule type: mRNA
A;Residues: 1-259 <ASH>
A;Cross-references: GB:M11565; NID:g205033; PIDN:AAA41466.1; PID:g205034
R;Lazure, C.; Leduc, R.; Seidah, N.G.; Thibault, G.; Genest, J.; Chretien, M.
Nature 307, 555-558, 1984
A;Title: Amino acid sequence of rat submaxillary tonin reveals similarities to serin
A;Reference number: A93323; MUID:84117504
A;Accession: A93323
A;Molecule type: protein
A;Residues: 25-259 <LA1>
R;Lazure, C.
submitted to the Protein Sequence Database, March 1985
A;Reference number: A94635
A;Contents: carbohydrate-binding site; revisions
A;Accession: A94635
A;Molecule type: protein
A;Residues: 104-119 <LA2>
R;Kamada, M.; Furuhashi, N.; Yamaguchi, T.; Ikekita, M.; Kizuki, K.; Moriya, H.
Biochem. Biophys. Res. Commun. 166, 231-237, 1990
A;Title: Observation of tissue prokallikrein activation by some serine proteases, ar
A;Reference number: A34050; MUID:90147705
A;Accession: A34050
A;Molecule type: protein
A;Residues: 25-30, 'X', 32-34 <RAM>
R;Lazure, C.; Leduc, R.; Seidah, N.G.; Thibault, G.; Genest, J.; Chretien, M.
Biochem. Cell Biol. 65, 321-337, 1987
A;Title: The complete amino acid sequence of rat submaxillary gland tonin does conta
A;Reference number: A30971; MUID:87271223
A;Accession: A30971
A;Molecule type: protein
A;Residues: 25-259 <LA2>
C;Comment: This protein is found in submaxillary gland. It has both trypsin- and chy
C;Genetics:
A;Introns: 16/1; 67/2; 163/1; 208/3
C;Superfamily: trypsin; trypsin homology
C;Keywords: glycoprotein; hydrolase; serine proteinase
F;1-17/Domain: signal sequence #status predicted <SIG>
F;18-24/Domain: propeptide #status predicted <PRO>
F;25-259/Product: tonin #status experimental <MAT>
F;25-251/Domain: trypsin homology <TRY>
F;31-171,48-64,150-217,182-196,207-232/Disulfide bonds: #status predicted
F;63,118,211/Active site: His, Asp, Ser #status predicted
F;106,189/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 74.0%; Score 37; DB 1; Length 259;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||||| I
Db 213 GPLVCRGTL 221

RESULT 23
D23863
tissue kallikrein (EC 3.4.21.35) S3 precursor - rat
N;Alternate names: glandular kallikrein; kallikrein-like enzyme pKLUK9; kininogenin;
C;Species: Rattus norvegicus (Norway rat)
C;Date: 22-Jul-1987 #sequence_revision 22-Jul-1987 #text_change 22-Jun-1999
C;Accession: D23863; A23710; B23710; S46212; S19302
R;Ashley, P.L.; Macdonald, R.J.
Biochemistry 24, 4512-4520, 1985
A;Title: Kallikrein-related mRNAs of the rat submaxillary gland: nucleotide sequence
A;Reference number: A23863; MUID:86051477
A;Accession: D23863
A;Molecule type: mRNA
A;Residues: 1-259 <ASH>

A:Cross-references: GB:M11566; NID:g205035; PIDN:AAA41467.1; PID:g205036
A:Experimental source: submaxillary gland
R:Yamaguchi, T.; Carretero, O.A.; Scicli, A.G.
J. Biol. Chem. 266, 5011-5017, 1991
A:Title: A novel serine protease with vasoconstrictor activity coded by the kallikrein gene
A:Reference number: A23710; MUID:91161590
A:Accession: A23710
A:Status: preliminary
A:Molecule type: protein
A:Residues: 25-30, 'X', 33-36 <VAF>
A:Experimental source: submandibular gland
A:Accession: B23710
A:Status: preliminary
A:Molecule type: protein
A:Residues: 112-123 <YAM>
A:Experimental source: submandibular gland
R:Schoyen, H.; Wassdal, I.; Toft, K.; Almendingen, M.; Berg, T.
Biochem. J. 302, 229-235, 1994
A:Title: Purification of enzymes of the kallikrein gene family (rk8 and rk9) from the rat
A:Reference number: S46211; MUID:94347106
A:Accession: S46212
A:Molecule type: protein
A:Residues: 25-62;112-149 <SCH>
A:Experimental source: prostate
R:Berg, T.; Schoyen, H.; Wassdal, I.; Hull, R.; Gerskowitch, V.P.; Toft, K.
Biochem. J. 281, 819-828, 1992
A:Title: Characterization of a new kallikrein-like enzyme (KLP-S3) of the rat submandibular gland
A:Reference number: S19302; MUID:92162030
A:Accession: S19302
A:Molecule type: protein
A:Residues: 25-30, 'X', 32-45, 'X', 47, 'X', 49, 'X', 51-53;112-130 <BER>
A:Experimental source: submandibular gland
A:Experimental source: submandibular gland
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:25-251/Domain: trypsin homology <TRY>
F:25-111/Product: tissue kallikrein S3 light chain #status experimental <KHC>
F:112-259/Product: tissue kallikrein S3 heavy chain #status experimental <KLC>
F:63,118,211/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 259;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 213 GPLICDGL 221

RESULT 24
A29746
tissue kallikrein (EC 3.4.21.35) mGK-22 precursor - mouse
N:Alternate names: epidermal growth factor-binding protein type A
C:Species: Mus musculus (house mouse)
C:Date: 08-Mar-1989 #sequence_revision 30-Jun-1989 #text_change 22-Jun-1999
C:Accession: A29746; I70039; I70038
R:Drinkwater, C.C.; Evans, B.A.; Richards, R.I.
Biochemistry 26, 6750-6756, 1987
A:Title: Mouse glandular kallikrein genes: identification and characterization of the genes
A:Reference number: A90522; MUID:88107594
A:Accession: A29746
A:Molecule type: DNA
A:Residues: 1-259 <DRI>
A:Experimental source: strain BALB/c, salivary gland
A:Note: the authors translated the codon GAC for residue 210 as Gly
R:Evans, B.A.; Drinkwater, C.C.; Richards, R.I.
J. Biol. Chem. 262, 8027-8034, 1987
A:Title: Mouse glandular kallikrein genes: Structure and partial sequence analysis of the genes
A:Reference number: I55260; MUID:87250386
A:Accession: I70039
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 70-120 <RES>

A:Cross-references: GB:M18618; NID:gl98556; PIDN:AAA39362.1; PID:gl98561

A:Accession: I70038
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 17-54 <PE2>
A:Cross-references: GB:M18598; NID:gl98555; PIDN:AAA39361.1; PID:gl98560
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:25-251/Domain: trypsin homology <TRY>
F:65,118,211/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 259;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 213 GPLICDGL 221

RESULT 25
I56559
neuropsin - mouse
C:Species: Mus musculus (house mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 21-Jul-2000
C:Accession: I56559
R:Chen, Z.L.; Yoshida, S.; Kato, K.; Momota, Y.; Suzuki, J.; Tanaka, T.; Ito, J.; Nils
J. Neurosci. 15, 5088-5097, 1995
A:Title: Expression and activity-dependent changes of a novel limbic-serine protease
A:Reference number: I56559; MUID:95348817
A:Accession: I56559
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-260 <RES>
A:Cross-references: GB:D30785; NID:gl648847; PIDN:BAA06451.1; PID:gl020091
C:Superfamily: trypsin; trypsin homology
F:33-252/Domain: trypsin homology <TRY>

Query Match 74.0%; Score 37; DB 2; Length 260;
Best Local Similarity 77.8%; Pred. No. 22;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 214 GPLVCDGML 222

RESULT 26
KQMS1
tissue kallikrein (EC 3.4.21.35) mGK-1 precursor, submandibular - mouse
N:Alternate names: glandular kallikrein; kininogenin
C:Species: Mus musculus (house mouse)
C:Date: 19-Feb-1984 #sequence_revision 19-Feb-1984 #text_change 24-Sep-1999
C:Accession: A00941
R:Mason, A.J.; Evans, B.A.; Cox, D.R.; Shine, J.; Richards, R.I.
Nature 303, 300-307, 1983
A:Title: Structure of mouse kallikrein gene family suggests a role in specific processes
A:Reference number: A00941; MUID:83219214
A:Accession: A00941
A:Molecule type: DNA
A:Residues: 1-261 <MAS>
A:Cross-references: GB:V00829; NID:g52775; PIDN:CAA24213.1; PID:g52776
A:Experimental source: Ouakenbush inbred strain
C:Comment: mGK-1 belongs to a family of 25 to 30 homologous kallikrein genes.
C:Comment: Tissue kallikreins cleave Met-Lys and Arg-Ser bonds in kininogen to release
C:Genetics:
A:Gene: mGK-1
A:Map position: 7
A:Introns: 16/1; 69/2; 165/1; 210/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; saliva; serine proteinase; submandibular gland;
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-24/Domain: activation peptide #status predicted <APT>
F:25-261/Product: tissue kallikrein, submandibular #status predicted <MPT>

F:25-253/Domain: trypsin homology <TRY>
F:31-173,50-66,152-219,184-198,209-234/Disulfide bonds: #status predicted
F:65,120,213/Active site: His, Asp, Ser #status predicted
F:102/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 74.0%; Score 37; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 215 GPLICDGV 223

RESULT 27

NGMSG

7S nerve growth factor gamma chain (EC 3.4.21.-) precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 18-Dec-1981 #sequence_revision 17-May-1985 #text_change 18-Jun-1999
C:Accession: A91005; A90949; A93510; A92341; A00942; A21093; A22705
R:Evans, B.A.; Richards, R.I.
EMBO J. 4, 133-138, 1985
A:Title: Genes for the alpha and gamma subunits of mouse nerve growth factor are contiguous
A:Reference number: A91005; MUID:85257431
A:Accession: A91005
A:Molecule type: DNA
A:Residues: 1-261 <EVA>
R:Ullrich, A.; Gray, A.; Wood, W.I.; Hayflick, J.; Seeburg, P.H.
DNA 3, 387-392, 1984

A:Title: Isolation of a cDNA clone coding for the gamma-subunit of mouse nerve growth factor
A:Reference number: A90949; MUID:85076169
A:Accession: A90949

A:Molecule type: mRNA
A:Residues: 1-261 <URL>
A:CROSS-references: GB:X01389; NID:g53373; PIDN:CAA25645.1; PID:g53374
R:Howles, P.N.; Dickinson, D.P.; DiCaprio, L.L.; Woodworth-Gutai, M.; Gross, K.W.
Nucleic Acids Res. 12, 2791-2805, 1984
A:Title: Use of a cDNA recombinant for the gamma-subunit of mouse nerve growth factor to
A:Reference number: A93510; MUID:84169573
A:Accession: A93510
A:Molecule type: mRNA
A:Residues: 127-202, 'E', 204-261 <HOW>

A:CROSS-references: GB:X00472; NID:g54260; PIDN:CAA25154.1; PID:g54261
A:Experimental source: inbred strain DBA/2J
R:Thomas, K.A.; Baglan, N.C.; Bradshaw, R.A.
J. Biol. Chem. 256, 9156-9166, 1981

A:Title: The amino acid sequence of the gamma-subunit of mouse submaxillary gland 7 S ne
A:Reference number: A92341; MUID:81284363
A:Accession: A92341

A:Molecule type: protein
A:Residues: 25-107,112-261 <THO>

A:Experimental source: outbred strain Swiss Webster
C:Comment: 7S nerve growth factor is composed of two alpha chains, a beta dimer composed
C:Comment: The active form of the gamma chain occurs naturally as combinations of either
C:Genetics:

A:Map position: 7
A:Introns: 16/1; 69/2; 165/1; 210/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; growth factor; hydrolase; serine proteinase; submandibular gla
F:1-18/Domain: signal sequence #status predicted <SIG>
F:25-253/Domain: trypsin homology <TRY>
F:25-107/Domain: segment B1 <GB1>
F:25-107,112-261/Product: nerve growth factor gamma chain (active form) #status experime
F:112-261/Domain: segment A <GAA>
F:112-164/Domain: segment C <GCC>
F:165-261/Domain: segment B2 <GB2>
F:31-173,50-66,152-219,184-198,209-234/Disulfide bonds: #status predicted
F:65,120,213/Active site: His, Asp, Ser #status predicted
F:102/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 74.0%; Score 37; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 215 GPLICDGV 223

RESULT 28

TRMSM5

tissue kallikrein (EC 3.4.21.35) mGK-5, submandibular precursor - mouse
N:Alternate names: glandular kallikrein
C:Species: Mus musculus (house mouse)
C:Date: 31-Mar-1993 #sequence_revision 31-Mar-1993 #text_change 18-Jun-1999
C:Accession: S06305; I70019
R:Drinkwater, C.C.; Richards, R.I.
Nucleic Acids Res. 15, 10052, 1987
A:Title: Sequence of the mouse glandular kallikrein gene, mGK-5.
A:Reference number: S06305; MUID:88096499
A:Accession: S06305
A:Molecule type: DNA
A:Residues: 1-261 <ORI>
R:Evans, B.A.; Drinkwater, C.C.; Richards, R.I.
J. Biol. Chem. 262, 8027-8034, 1987
A:Title: Mouse glandular kallikrein genes: Structure and partial sequence analysis o
A:Reference number: I55260; MUID:87250386
A:Accession: I70019

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 70-122 <RES>

A:CROSS-references: GB:M18604; NID:g198514; PIDN:AA015284.1; PID:g198524

C:Genetics:

A:Introns: 16/1; 69/2; 165/1; 210/3

C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase; submandibular gland
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-24/Domain: propeptide #status predicted <PRO>
F:25-261/Product: tissue kallikrein mGK-5, submandibular #status predicted <MAT>
F:25-253/Domain: trypsin homology <TRY>
F:31-173,50-66,152-219,184-198,209-234/Disulfide bonds: #status predicted
F:65,120,213/Active site: His, Asp, Ser #status predicted
F:102/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 74.0%; Score 37; DB 1; Length 261;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||:| | |
Db 215 GPLICDGV 223

RESULT 29

A34079

tissue kallikrein (EC 3.4.21.35) p1 precursor - rat
N:Alternate names: kallikrein-related proteinase k8
C:Species: Rattus norvegicus (Norway rat)
C:Date: 30-Mar-1990 #sequence_revision 30-Mar-1990 #text_change 22-Jun-1999
C:Accession: A34079; S10700
R:Brady, J.M.; Wines, D.R.; MacDonald, R.J.
Biochemistry 28, 5203-5210, 1989
A:Title: Expression of two kallikrein gene family members in the rat prostate.
A:Reference number: A34079; MUID:89352606

A:Accession: A34079

A:Status: preliminary

A:Molecule type: DNA; mRNA

A:Residues: 1-261 <BRA>

A:CROSS-references: GB:M27215; GB:M27216; GB:M27217; NID:g206638; PIDN:AAA42036.1; P

A:Experimental source: prostate

R:Elmoujahed, A.; Gutman, N.; Brillard, M.; Gauthier, F.

FEBS Lett. 265, 137-140, 1990

A:Title: Substrate specificity of two kallikrein family gene products isolated from

A:Reference number: S10698; MUID:90306305

A:Accession: S10700

A:Molecule type: protein
A:Residues: 25-43;112-138 <ELM>
A:Experimental source: submaxillary gland
A:Note: 125-Lys was also found
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:25-253/Domain: trypsin homology <TRY>
F:25-111/Product: tissue kallikrein p1
F:112-261/Product: tissue kallikrein p1 light chain #status experimental <MAT1>
F:112-261/Product: tissue kallikrein p1 heavy chain #status experimental <MAT2>
F:65,120,213/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 261;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
| | | : | | |
Db 215 GPLICDGV 223

RESULT 30

tissue kallikrein (EC 3.4.21.35) precursor - mouse
N:Alternate names: glandular kallikrein
C:Species: Mus musculus (house mouse)
C>Date: 01-Dec-1989 #sequence_revision 01-Dec-1989 #text_change 21-Jul-2000
C:Accession: S01971; I70023
R:Drinkwater, C.C.; Richards, R.I.
Nucleic Acids Res. 15, 10918, 1988
A:Title: Sequence of mGK-11, a mouse glandular kallikrein gene.
A:Reference number: S01971; MUID:89083511
A:Accession: S01971
A:Molecule type: DNA
A:Residues: 1-261 <ORI>

A:Cross-references: EMBL:X13215; NID:g53064; PIDN:CAA31604.3; PID:g5708211
J. Biol. Chem. 262, 8027-8034, 1987
A:Title: Mouse glandular kallikrein genes: Structure and partial sequence analysis of the
A:Reference number: I55260; MUID:87250386
A:Accession: I70023
A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA
A:Residues: 70-122 <RES>
A:Cross-references: GB:M18610; NID:g198529; PIDN:AAA39353.1; PID:g198537
C:Genetics:
A:Introns: 16/1; 69/2; 165/1; 210/3

C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-17/Domain: signal sequence #status predicted <SIG>
F:18-24/Domain: propeptide #status predicted <PRO>
F:25-261/Product: tissue kallikrein #status predicted <MAT>
F:25-253/Domain: trypsin homology <TRY>
F:65,120,213/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 261;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
| | | : | | |
Db 215 GPLICDGV 223

RESULT 31

gamma-tenin (EC 3.4.21.54) precursor - mouse
C:Species: Mus musculus (house mouse)
C>Date: 19-May-1989 #sequence_revision 19-May-1989 #text_change 22-Jun-1999
C:Accession: A28062; I70033; I70032
R:Drinkwater, C.C.; Evans, B.A.; Richards, R.I.
J. Biol. Chem. 263, 8565-8568, 1988
A:Title: Sequence and expression of mouse gamma-renin.
A:Reference number: A28062; MUID:88243703

A:Accession: A28062
A:Molecule type: mRNA
A:Residues: 1-261 <ORI>
A:Cross-references: GB:J03877; NID:g200699; PIDN:AAA40049.1; PID:g200700
J. Biol. Chem. 262, 8027-8034, 1987
A:Title: Mouse glandular kallikrein genes: Structure and partial sequence analysis of
A:Reference number: I55260; MUID:87250386
A:Accession: I70033
A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA
A:Residues: 70-122 <RES>
A:Cross-references: GB:M18615; NID:g198546; PIDN:AAA39358.1; PID:g198549
A:Accession: I70032
A:Status: preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA
A:Residues: 17-54 <RE2>
A:Cross-references: GB:M18594; NID:g198545; PIDN:AAA39357.1; PID:g198548
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-261/Product: gamma-renin #status predicted <MAT>
F:25-253/Domain: trypsin homology <TRY>
F:65,120,213/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 261;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
| | | : | | |
Db 215 GPLICDGV 223

RESULT 32

A25606
tissue kallikrein (EC 3.4.21.35) submandibular precursor - mouse
N:Alternate names: glandular kallikrein; kininogenin
C:Species: Mus musculus (house mouse)
C>Date: 19-Nov-1988 #sequence_revision 30-Jun-1991 #text_change 13-Nov-1998
C:Accession: A25606; S06661
R:van Leeuwen, B.H.; Evans, B.A.; Tregear, G.W.; Richards, R.I.
J. Biol. Chem. 261, 5529-5535, 1986
A:Title: Mouse glandular kallikrein genes. Identification, structure, and expression
A:Reference number: A25606; MUID:86168299
A:Accession: A25606

A:Molecule type: DNA
A:Residues: 1-261 <VAN>
A:Cross-references: GB:M13498
A:Experimental source: renal
A:Note: the authors translated the codon GTC for residue 57 as Ala, CTC for residue 1
R:Murakami, K.; Ikigai, H.; Nagumo, N.; Tomita, M.; Shimamura, T.
FEBS Lett. 257, 400-402, 1989
A:Title: A cytosolic tissue kallikrein isolated from mouse submandibular glands.
A:Reference number: S06660; MUID:90060365
A:Accession: S06661
A:Molecule type: protein
A:Residues: 165-174 <MUR>
C:Genetics:
A:Gene: mGK-6

A:Introns: 16/1; 69/2; 165/1; 210/3
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-261/Product: tissue kallikrein #status predicted <MAT>
F:25-253/Domain: trypsin homology <TRY>
F:65,120,213/Active site: His, Asp, Ser #status predicted

Query Match 74.0%; Score 37; DB 2; Length 261;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

```
Db 215 GPLICDGV 223
|||||
RESULT 33
A29745
tissue kallikrein (EC 3.4.21.35) mGK-9 precursor, submandibular - mouse
N:Alternate names: glandular kallikrein mGK-9; major epidermal growth factor-binding protein
C:Species: Mus musculus (house mouse)
C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 22-Jun-1999
C:Accession: C29746; A29745; A27120; I70015
R:Drinkwater, C.C.; Evans, B.A.; Richards, R.I.
Biochemistry 26, 6750-6756, 1987
A:Title: Mouse glandular kallikrein genes: identification and characterization of the genes
A:Reference number: A90522; MUID:88107594
A:Accession: C29746
A:Molecule type: DNA
A:Residues: 1-261 <DRI>
A:Cross-references: GB:M17985; NID:g193476; PIDN:AAA37681.1; PID:g387166
A:Experimental source: strain BALB/C, salivary gland
R:Blaber, M.; Isackson, P.J.; Bradshaw, R.A.
Biochemistry 26, 6742-6749, 1987
A:Title: A complete cDNA sequence for the major epidermal growth factor binding protein
A:Reference number: A29745; MUID:88107593
A:Accession: A29745
A:Molecule type: mRNA
A:Residues: 1-261 <BLA>
A:CROSS-references: GB:M17962; NID:g192997; PIDN:AAA37541.1; PID:g309212
R:Isackson, P.J.; Silverman, R.E.; Blaber, M.; Server, A.C.; Nichols, R.A.; Shooter, E.M.
Biochemistry 26, 2082-2085, 1987
A:Title: Epidermal growth factor binding protein: identification of a different protein.
A:Reference number: A27120; MUID:87299636
A:Accession: A27120
A:Molecule type: protein
A:Residues: 25-54;112-124, 'X',126-130;165-184,'X',186-187,'X',189-192 <ISA>
R:Evans, B.A.; Drinkwater, C.C.; Richards, R.I.
J. Biol. Chem. 262, 8027-8034, 1987
A:Title: Mouse glandular kallikrein genes: Structure and partial sequence analysis of the genes
A:Reference number: I55260; MUID:87250386
A:Accession: I70015
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 70-122 <RES>
A:CROSS-references: GB:M18608; NID:g198500; PIDN:AAA39351.1; PID:g198506
C:Comment: This sequence is one of approximately twenty-five members of a gene family on chromosome 12p11.
C:Genetics:
A:Gene: KAL
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase; zymogen
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-24/Domain: propeptide #status predicted <PRO>
F:25-261/Product: tissue kallikrein mGK-9, submandibular #status experimental <MAT>
F:25-233/Domain: trypsin homology <TRY>
F:65,120,213/Active site: His, Asp, Ser #status predicted
Query Match 74.0%; Score 37; DB 2; Length 261;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 215 GPLICDGV 223

RESULT 34
S45303
tissue kallikrein (EC 3.4.21.35) precursor - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 20-Oct-1994 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C:Accession: S45303; S38487
R:Gauthier, E.R.; Dumas, C.; Chapdelain, P.; Tremblay, R.R.; Dube, J.Y.
Biochim. Biophys. Acta 1218, 102-104, 1994
A:Title: Characterization of canine pancreas kallikrein cDNA.
A:Reference number: A92823; MUID:89009866
A:Accession: C30525
```

```
A:Reference number: S45303; MUID:94250683
A:Accession: S45303
A:Molecule type: mRNA
A:Residues: 1-261 <GAU>
A:CROSS-references: EMBL:X75479; NID:g414018; PIDN:CAA53210.1; PID:g414019
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-261/Product: tissue kallikrein #status predicted <MAT>
F:25-233/Domain: trypsin homology <TRY>
F:65,120,213/Active site: His, Asp, Ser #status predicted
Query Match 74.0%; Score 37; DB 2; Length 261;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 215 GPLICDGV 223

RESULT 35
S15686
tissue kallikrein (EC 3.4.21.35) precursor, renal - multimammate rat (Mastomys natalensis)
C:Species: Mastomys natalensis
C:Date: 09-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 22-Jun-1999
C:Accession: I83227; S15686
R:Fahnestock, M.
DNA Cell Biol. 13, 293-300, 1994
A:Title: Characterization of kallikrein cDNAs from the African rodent Mastomys.
A:Reference number: I60208; MUID:94226702
A:Accession: I83227
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-263 <RES>
A:CROSS-references: EMBL:X17352; NID:g55526; PIDN:CAA35232.1; PID:g55527
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:25-255/Domain: trypsin homology <TRY>
F:65,121,215/Active site: His, Asp, Ser #status predicted
Query Match 74.0%; Score 37; DB 2; Length 263;
Best Local Similarity 66.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 217 GPLICDGV 225

RESULT 36
S65663
granzyme 3 (EC 3.4.21.-) precursor - human
N:Alternate names: cytotoxic T-lymphocyte proteinase 3
C:Species: Homo sapiens (man)
C:Date: 06-Dec-1996 #sequence_revision 13-Mar-1997 #text_change 22-Jun-1999
C:Accession: S65663; S65617; C30525
R:Przetak, M.M.; Roast, S.; Schmidt, B.F.
FEBS Lett. 364, 268-271, 1995
A:Title: Cloning of cDNA for human granzyme 3.
A:Reference number: S65617; MUID:95278340
A:Accession: S65663
A:Molecule type: mRNA
A:Residues: 1-264 <PRZ>
A:CROSS-references: EMBL:U26174; NID:g829637; PIDN:AAA74578.1; PID:g829638
A:Accession: S65617
A:Molecule type: protein
A:Residues: 27-42 <PRW>
R:Hameed, A.; Lowrey, D.M.; Lichtenheld, M.; Podack, E.R.
J. Immunol. 141, 3142-3147, 1988
A:Title: Characterization of three serine esterases isolated from human IL-2 activated T cells.
A:Reference number: A92823; MUID:89009866
A:Accession: C30525
```

A:Molecule type: protein
A:Residues: 27-42 <HAM>
C:Genetics:

A:Gene: GDB:G2MK: PRSS
A:Cross-references: GDB:605612; OMIM:600784
A:Map position: 14q11.2-14q11.2
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-26/Domain: signal sequence #status predicted <SIG>
F:27-264/Product: granzyme 3 #status experimental <MAT>
F:27-254/Domain: trypsin homology <TRY>

Query Match 74.0%; Score 37; DB 2; Length 264;
Best Local Similarity 71.4%; Pred. No. 22;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
|||||
Db 216 GPLCKG 222

RESULT 37

T12994

riboflavin biosynthesis protein ribG homolog T21L8.140 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 26-May-2000

C:Accession: T12994

R:Choisne, N.; Robert, C.; Brottier, P.; Wincker, P.; Cattolico, L.; Artiguenave, F.; Sa

submitted to the Protein Sequence Database, July 1999

A:Reference number: 217586

A:Accession: T12994

A:Molecule type: DNA

A:Residues: 1-599 <CHO>

A:Cross-references: EMBL:AL096860; GSPDB:GN00061; ATSP:T21L8.140

A:Experimental source: cultivar Columbia; BAC clone T21L8

C:Genetics:

A:Gene: ATSP:T21L8.140

A:Map position: 3

A:Introns: 119/3; 164/3; 198/1; 240/1; 344/3; 384/3; 437/1; 513/3

Query Match 74.0%; Score 37; DB 2; Length 599;
Best Local Similarity 75.0%; Pred. No. 46;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 PLVCRGTL 9
|||||
Db 12 PLICRATL 19

RESULT 38

I52972

kallikrein - mouse (fragment)

C:Species: Mus musculus (house mouse)

C>Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 22-Jun-1999

C:Accession: I52972

R:Nordeen, S.K.; Mason, A.J.; Richards, R.I.; Baxter, J.D.; Shine, J.

DNA 1, 309-311, 1982

A:Title: Mouse kallikrein arginyl-esteropeptidase genes: Analysis of cloned cDNAs sugges

A:Reference number: I52972; MUID:83182015

A:Accession: I52972

A>Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-66 <RES>

A:Cross-references: GB:K01654; NID:g198490; PIDN:AAA39346.1; PID:g198491

C:Superfamily: trypsin; trypsin homology

F:1-58/Domain: trypsin homology (fragment) <TRY>

Query Match 72.0%; Score 36; DB 2; Length 66;
Best Local Similarity 66.7%; Pred. No. 9;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||

Db 20 GPLICDGIL 28

RESULT 39

S44462

elastase (EC 3.4.21.-) 2B - horse (fragments)

C:Species: Equus caballus (domestic horse)

C>Date: 20-Oct-1994 #sequence_revision 01-Nov-1996 #text_change 07-May-1999

C:Accession: S44462

R:Dubin, A.; Potempa, J.; Travis, J.

Biochem. J. 300, 401-406, 1994

A:Title: Structural and functional characterization of elastases from horse neutrophil

A:Reference number: S44461; MUID:94271153

A:Accession: S44462

A:Molecule type: protein

A:Residues: 1-31;32-56;57-73 <DUB>

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; serine proteinase

F:1-73/Domain: trypsin homology (fragments) <TRY>

Query Match 72.0%; Score 36; DB 2; Length 73;
Best Local Similarity 85.7%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
|||||
Db 66 GPLVCNG 72

RESULT 40

K0M5M

tissue kallikrein (EC 3.4.21.35), submandibular - mouse (fragment)

N:Alternate names: glandular kallikrein; kininogenin

C:Species: Mus musculus (house mouse)

C>Date: 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change 16-Jun-2000

C:Accession: A00939

R:Richards, R.I.; Catanzaro, D.F.; Mason, A.J.; Morris, B.J.; Baxter, J.D.; Shine, J.

J. Biol. Chem. 257, 2758-2761, 1982

A:Title: Mouse glandular kallikrein genes. Nucleotide sequence of cloned cDNA coding

A:Reference number: A00939; MUID:82142394

A:Accession: A00939

A:Molecule type: mRNA

A:Residues: 1-149 <RIC>

A:Cross-references: GB:V00828; GB:J00389; NID:g52773; PIDN:CAA24211.1; PID:g1334112

A:Experimental source: Quakenbush inbred strain

C:Comment: Tissue kallikreins cleave Met-Lys and Arg-Ser bonds in kininogen to releas

C:Genetics:

A:Map position: 7

C:Superfamily: trypsin; trypsin homology

C:Keywords: hydrolase; saliva; serine proteinase; submandibular gland

F:1-141/Domain: trypsin homology (fragment) <TRY>

F:8,101/Active site: Asp, Ser #status predicted

F:40-107,72-86,97-122/Disulfide bonds: #status predicted

Query Match 72.0%; Score 36; DB 1; Length 149;
Best Local Similarity 66.7%; Pred. No. 20;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 103 GPLICDGIL 111

RESULT 41

S69370

duodenase - bovine

C:Species: Bos primigenius taurus (cattle)

C>Date: 24-Jul-1998 #sequence_revision 24-Jul-1998 #text_change 21-Nov-1998

C:Accession: S69370

R:zamolodchikova, T.S.; Vorotyntseva, T.I.; Nazimov, I.V.; Grishina, G.A.

Eur. J. Biochem. 227, 873-879, 1995

A:Title: Duodenase, a new serine protease of unusual specificity from bovine duodenal

A:Reference number: S69370; MUID:95172076

A:Accession: S69370
A:Molecule type: protein
A:Residues: 1-226 <ZAM>
C:Superfamily: trypsin; trypsin homology
F:1-218/Domain: trypsin homology <TRY>

Query Match 72.0%; Score 36; DB 2; Length 226;
Best Local Similarity 85.7%; Pred. No. 29;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRG 7
||| ||| |
Db 183 GPLVCG 189

RESULT 42

TRPSN
trypsin (EC 3.4.21.4) precursor - spiny dogfish
N:Alternate names: trypsinogen
C:Species: Squalus acanthias (spiny dogfish)
C:Date: 31-May-1979 #sequence_revision 31-May-1979 #text_change 21-Jan-1997
C:Accession: A00950; B27719
R:Titanil, K.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
Biochemistry 14, 1358-1366, 1975
A:Title: Amino acid sequence of dogfish trypsin.
A:Reference number: A00950; MUID:75146445
A:Accession: A00950
A:Molecule type: protein
A:Residues: 8-229 <TIT>
A:Note: 119-Pro was also found
R:Hermoldson, M.A.; Tye, R.W.; Reeck, G.R.; Neurath, H.; Walsh, K.A.
FEBS Lett. 14, 222-224, 1971
A:Title: Comparison of the amino terminal sequences of bovine, dogfish, and lungfish try
A:Reference number: A27719
A:Accession: B27719
A:Molecule type: protein
A:Residues: 1-21 <HER>
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; polymorphism; protein digestion; serine proteinase; zym
F:1-7/Domain: activation peptide #status experimental <APT>
F:8-229/Domain: trypsin #status predicted <MAT>
F:14-143,32-48,116-123,189,154-168,179-203/Disulfide bonds: #status predicted
F:47,91,183/Active site: His, Asp, Ser #status predicted
F:59,61,64,69/Binding site: calcium (Glu, Asp, Ala, Glu) #status predicted

Query Match 72.0%; Score 36; DB 1; Length 229;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||| ||| |
Db 185 GPVVCNGML 193

RESULT 43

TRPCTR
trypsin (EC 3.4.21.4) precursor - pig (tentative sequence)
N:Contains: trypsinogen
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 24-Apr-1984 #sequence_revision 24-Apr-1984 #text_change 31-Mar-2000
C:Accession: A90641; A90368; A00947
R:Charles, M.; Rovey, M.; Guidoni, A.; Desnuelle, P.
Biochim. Biophys. Acta 69, 115-129, 1963
A:Title: Su le trypsinogene et la trypsine de porc.
A:Reference number: A90641
A:Accession: A90641
A:Molecule type: protein
A:Residues: 1-10 <CHA>
R:Hermoldson, M.A.; Ericsson, L.H.; Neurath, H.; Walsh, K.A.
Biochemistry 12, 3146-3153, 1973
A:Title: Determination of the amino acid sequence of porcine trypsin by sequenator analy
A:Reference number: A90368; MUID:73258692

A:Accession: A90368
A:Molecule type: protein
A:Residues: 9-231 <HER>
A:Note: at position 20, Ile and Val occur alternatively
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; pancreas; polymorphism; protein digestion; serine proteinase;
F:1-231/Product: trypsinogen #status experimental <ZYM>
F:1-8/Domain: activation peptide #status experimental <APT>
F:9-231/Product: trypsin #status experimental <MAT>
F:9-224/Domain: trypsin homology <TRY>
F:15-145,33-49,117-218,124-191,156-170,181-205/Disulfide bonds: #status predicted
F:48,92,185/Active site: His, Asp, Ser #status predicted
F:60,62,65,70/Binding site: calcium (Glu, Asn, Val, Glu) #status predicted

Query Match 72.0%; Score 36; DB 1; Length 231;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||| ||| |
Db 187 GPVVCNGQL 195

RESULT 44

S31778
trypsin (EC 3.4.21.4) II precursor - Atlantic salmon (fragment)
C:Species: Salmo salar (Atlantic salmon)
C:Date: 22-Nov-1993 #sequence_revision 03-Aug-1995 #text_change 22-Jun-1999
C:Accession: S66658; S31778
R:Male, R.; Lorens, J.B.; Smalas, A.O.; Torrisen, K.R.
Eur. J. Biochem. 232, 677-685, 1995
A:Title: Molecular cloning and characterization of anionic and cationic variants of
A:Reference number: S66657; MUID:96035908
A:Accession: S66658
A:Molecule type: mRNA
A:Residues: 1-231 <MAL>
A:Cross-references: EMBL:X70073; NID:g64385; PIDN:CAA49678.1; PID:g64386
C:Superfamily: trypsin; trypsin homology
C:Keywords: hydrolase; serine proteinase
F:1-4/Domain: signal sequence (fragment) #status predicted <SIG>
F:5-9/Domain: activation peptide #status predicted <APT>
F:10-231/Product: trypsin II #status predicted <MAT>
F:10-224/Domain: trypsin homology <TRY>
F:16-145,34-50,118-218,125-191,156-170,181-205/Disulfide bonds: #status predicted
F:49,93,185/Active site: His, Asp, Ser #status predicted

Query Match 72.0%; Score 36; DB 2; Length 231;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||| ||| |
Db 187 GPVVCNGEL 195

RESULT 45

A32121
snake venom factor V activator (EC 3.4.21.95) alpha [validated] - Russell's viper
N:Alternate names: factor V-activating proteinase alpha; proteinase RVV-V alpha
C:Species: Vipera russellii siamensis
C:Date: 19-May-2000 #sequence_revision 19-May-2000 #text_change 18-Jul-2001
C:Accession: A32121
R:Tokunaga, F.; Nagasawa, K.; Tamura, S.; Miyata, T.; Iwanaga, S.; Kiesel, W.
J. Biol. Chem. 263, 17471-17481, 1988
A:Title: The factor V-activating enzyme (RVV-V) from Russell's viper venom. Identifi
A:Reference number: A94684; MUID:89034273
A:Accession: A32121
A:Molecule type: protein
A:Residues: 1-236 <YOK>
C:Comment: This enzyme can fully activate human factor V by a single cleavage.
C:Superfamily: trypsin; trypsin homology
C:Keywords: glycoprotein; hydrolase; serine proteinase
F:1-222/Domain: trypsin homology <TRY>

F:7-141,28-44,76-234,120-188,152-167,178-203/Disulfide bonds: #status predicted
 F:43,88,182/Active site: His, Asp, Ser #status predicted
 F:229/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 72.0%; Score 36; DB 1; Length 236;
 Best Local Similarity 55.8%; Pred. No. 30;
 Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 |||:
 Db 184 GPLICNGQI 192

Search completed: November 6, 2002, 12:08:19
 Job time : 12 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 8.11111 Seconds
(without alignments)
27.102 Million cell updates/sec

Title: US-09-905-083-80
Perfect score: 50
Sequence: 1 GPLVCRGTL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/iaa/PCRUS_COMB.pep.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	50	100.0	9	US-09-502-600-80	Sequence 80, Appl
2	50	100.0	224	US-08-944-483-33	Sequence 33, Appl
3	50	100.0	225	US-08-557-146-12	Sequence 12, Appl
4	50	100.0	225	US-09-027-337-4	Sequence 4, Appl
5	50	100.0	225	US-09-154-344-12	Sequence 12, Appl
6	50	100.0	253	US-08-557-146-2	Sequence 2, Appl
7	50	100.0	253	US-08-824-874-3	Sequence 3, Appl
8	50	100.0	253	US-09-154-344-2	Sequence 2, Appl
9	50	100.0	253	US-08-930-188-2	Sequence 2, Appl
10	50	100.0	253	US-09-210-084-3	Sequence 3, Appl
11	50	100.0	253	PCT-US96-04294-2	Sequence 2, Appl
12	40	80.0	20	US-08-472-228A-19	Sequence 19, Appl
13	40	80.0	20	US-09-146-831-19	Sequence 19, Appl
14	40	80.0	20	PCT-US96-09303-19	Sequence 19, Appl
15	40	80.0	237	US-08-096-946-10	Sequence 10, Appl
16	40	80.0	237	US-08-096-946-11	Sequence 11, Appl
17	40	80.0	237	US-08-844-024-2	Sequence 2, Appl
18	40	80.0	237	US-08-718-547-2	Sequence 2, Appl
19	40	80.0	237	US-08-768-859A-1	Sequence 1, Appl
20	40	80.0	237	US-08-768-859A-16	Sequence 16, Appl
21	40	80.0	237	US-08-768-859A-21	Sequence 21, Appl
22	40	80.0	237	US-08-767-820A-1	Sequence 1, Appl
23	40	80.0	237	US-08-767-820A-16	Sequence 16, Appl
24	40	80.0	237	US-08-767-820A-21	Sequence 21, Appl
25	40	80.0	237	US-08-622-046B-1	Sequence 1, Appl
26	40	80.0	237	US-08-622-046B-7	Sequence 7, Appl
27	40	80.0	237	US-08-622-046B-12	Sequence 12, Appl

28	40	80.0	237	4	US-08-944-483-37	Sequence 37, Appl
29	40	80.0	237	4	US-08-944-483-38	Sequence 38, Appl
30	40	80.0	237	4	US-09-100-264-1	Sequence 1, Appl
31	40	80.0	237	4	US-09-100-264-3	Sequence 3, Appl
32	40	80.0	237	4	US-09-100-264-12	Sequence 12, Appl
33	40	80.0	237	5	PCT-US94-07329-10	Sequence 10, Appl
34	40	80.0	237	5	PCT-US94-07329-11	Sequence 11, Appl
35	40	80.0	237	5	PCT-US95-06157-1	Sequence 1, Appl
36	40	80.0	237	5	PCT-US95-06157-16	Sequence 16, Appl
37	40	80.0	238	3	US-08-768-859A-8	Sequence 8, Appl
38	40	80.0	238	3	US-08-767-820A-8	Sequence 8, Appl
39	40	80.0	238	5	PCT-US95-06157-8	Sequence 8, Appl
40	40	80.0	240	1	US-08-472-228A-1	Sequence 1, Appl
41	40	80.0	240	4	US-09-146-831-1	Sequence 1, Appl
42	40	80.0	240	5	PCT-US96-09303-1	Sequence 1, Appl
43	40	80.0	244	3	US-08-768-859A-10	Sequence 10, Appl
44	40	80.0	244	3	US-08-767-820A-10	Sequence 10, Appl
45	40	80.0	244	3	US-08-622-046B-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-09-502-600-80
; Sequence 80, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 80
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 207-215 of the SCCE protein
US-09-502-600-80

Query Match 100.0%; Score 50; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
Db 1 GPLVCRGTL 9
|||||

RESULT 2
US-08-944-483-33
; Sequence 33, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KASS, MICHAEL R.
; APPLICANT: STEWART, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories

```
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
; US-08-944-483-33

Query Match 100.0%; Score 50; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
Db 178 GPLVCRGTL 186

RESULT 3
US-08-557-146-12
; Sequence 12, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181

; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
; US-08-944-483-33

Query Match 100.0%; Score 50; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
Db 178 GPLVCRGTL 186

RESULT 3
US-08-557-146-12
; Sequence 12, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
; US-08-557-146-12

Query Match 100.0%; Score 50; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
Db 179 GPLVCRGTL 187

RESULT 4
US-09-027-337-4
; Sequence 4, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hirokoshi
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of Scce homologous to
; OTHER INFORMATION: similar domain in TADG-15
; US-09-027-337-4

Query Match 100.0%; Score 50; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.079;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
Db 179 GPLVCRGTL 187

RESULT 5
US-09-154-344-12
; Sequence 12, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/154,344
FILING DATE: 16-SEP-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/557,146
FILING DATE: 14-DEC-1995
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-181
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 225 amino acids
TYPE: amino acids
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: polypeptide
US-09-154-344-12

Query Match 100.0%; Score 50; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.079; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 179 GPLVCRGTL 187

RESULT 6

US-08-557-146-2
Sequence 2, Application US/08557146
Patent No. 5834290
GENERAL INFORMATION:
APPLICANT: Egelrud, Torbjorn
APPLICANT: Hansson, Lennart
TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
TITLE OF INVENTION: Enzyme (SCCE)
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: White & Case, Patent Department
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2787
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/557,146
FILING DATE: 14-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sterner, Richard J.
REGISTRATION NUMBER: 35,372
REFERENCE/DOCKET NUMBER: 1103326-181
TELEPHONE: (212) 819-8783
TELEFAX: (212) 354-8113
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-557-146-2

Query Match 100.0%; Score 50; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.089; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 207 GPLVCRGTL 215

RESULT 7

US-08-824-874-3
Sequence 3, Application US/08824874
Patent No. 5962300
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL KALLIKREIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/824,874
FILING DATE: Filed Herewith
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0252 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 532504
US-08-824-874-3

Query Match 100.0%; Score 50; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.089;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 207 GPLVCRGTL 215

RESULT 8

US-09-154-344-2
Sequence 2, Application US/09154344
Patent No. 5981256
GENERAL INFORMATION:
APPLICANT: Egelrud, Torbjorn
APPLICANT: Hansson, Lennart

```
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (BPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-154-344-2

Query Match 100.0%; Score 50; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.089;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 207 GPLVCRGTL 215

RESULT 9
US-09-930-188-2
; Sequence 2, Application US/08930188
; Patent No. 6093397
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,188
; FILING DATE:
```

```
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-930-188-2

Query Match 100.0%; Score 50; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.089;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 207 GPLVCRGTL 215

RESULT 10
US-09-210-084-3
; Sequence 3, Application US/09210084
; Patent No. 6197511
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/210,084
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: Genbank
; CLONE: 532504
```

US-09-210-084-3

Query Match 100.0%; Score 50; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.089; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 207 GPLVCRGTL 215

RESULT 11
PCT-US96-04294-2
; Sequence 2, Application PC/TUS9604294
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
; TITLE OF INVENTION: RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/04294
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Bialock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090
; TELEFAX: 317-276-3861
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US96-04294-2

Query Match 100.0%; Score 50; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.089; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 207 GPLVCRGTL 215

RESULT 12
US-08-472-228A-19
; Sequence 19, Application US/08472228A
; Patent No. 5807978
; GENERAL INFORMATION:
; APPLICANT: Kokolus, William J.
; APPLICANT: Fritzsche, Herbert A.
; APPLICANT: Johnston, Dennis A.

; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES OF PROSTATE
; TITLE OF INVENTION: SPECIFIC ANTIGEN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,228A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSC:388
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-472-228A-19

Query Match 80.0%; Score 40; DB 1; Length 20;
Best Local Similarity 77.8%; Pred. No. 0.44;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 10 GPLVCGVL 18

RESULT 13
US-09-146-831-19
; Sequence 19, Application US/09146831
; Patent No. 6326471
; GENERAL INFORMATION:
; APPLICANT: Kokolus, William J.
; APPLICANT: Fritzsche, Herbert A.
; APPLICANT: Johnston, Dennis A.
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES OF PROSTATE
; TITLE OF INVENTION: SPECIFIC ANTIGEN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,831
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/472,228
FILING DATE: 07-JUN-1995
ATTORNEY/AGENT INFORMATION:
NAME: Highlander, Steven L.
REGISTRATION NUMBER: 37,642
REFERENCE/DOCKET NUMBER: UTSC:388
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
TELEX: 79-0924
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-09-146-831-19

Query Match 80.0%; Score 40; DB 4; Length 20;
Best Local Similarity 77.8%; Pred. No. 0.44;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 10 GPLVCNGVL 18

RESULT 14
PCT-US96-09303-19
Sequence 19, Application PC/TUS9609303
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: IMMUNOGENIC PEPTIDES OF PROSTATE
SPECIFIC ANTIGEN
NUMBER OF SEQUENCES: 22
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/09303
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/472,228
FILING DATE: 07-JUN-1995
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 20 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
PCT-US96-09303-19

Query Match 80.0%; Score 40; DB 5; Length 20;
Best Local Similarity 77.8%; Pred. No. 0.44;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 10 GPLVCNGVL 18

RESULT 15
US-08-096-946-10
Sequence 10, Application US/08096946
Patent No. 5516639
GENERAL INFORMATION:
APPLICANT: Tindall, Donald J.
APPLICANT: Young, Charles Y-F

TITLE OF INVENTION: Antibodies Specific for Human
TITLE OF INVENTION: Prostate Glandular Kallikrein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 NO. 5516639west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,946
FILING DATE: 19930722
CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 1016.62-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-8081
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
CLONE: Amino acid sequence of hK2
CLONE: (deduced from cDNA sequence)
US-08-096-946-10
Query Match 80.0%; Score 40; DB 1; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
|||||
Db 191 GPLVCNGVL 199
RESULT 16
US-08-096-946-11
Sequence 11, Application US/08096946
Patent No. 5516639
GENERAL INFORMATION:
APPLICANT: Tindall, Donald J.
APPLICANT: Young, Charles Y-F
APPLICANT: Klee, George G.
TITLE OF INVENTION: Antibodies Specific for Human
TITLE OF INVENTION: Prostate Glandular Kallikrein
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Merchant & Gould
STREET: 3100 NO. 5516639west Center
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/096,946
FILING DATE: 19930722

CLASSIFICATION: 436
ATTORNEY/AGENT INFORMATION:
NAME: Woessner, Warren D.
REGISTRATION NUMBER: 30,440
REFERENCE/DOCKET NUMBER: 1016.62-US-01
TELECOMMUNICATION INFORMATION:
TELEPHONE: 612-332-5300
TELEFAX: 612-332-8081
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
ORIGINAL SOURCE:
ORGANISM: Amino acid sequence of PSA (hk3)
US-08-096-946-11

Query Match 80.0%; Score 40; DB 1; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
DB 191 GPLVCNGVL 199

RESULT 17
US-08-844-024-2
Sequence 2, Application US/08844024
Patent No. 5840494
GENERAL INFORMATION:
APPLICANT: Katz, Aaron E., et al.
TITLE OF INVENTION: A Method for Molecular Staging Of
TITLE OF INVENTION: Prostate Cancer
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/844,024
FILING DATE: 18-APR-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/229,391
FILING DATE: 15-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 43677/JPW/TEP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOP UI
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-844-024-2

Query Match 80.0%; Score 40; DB 2; Length 237;

Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
DB 191 GPLVCNGVL 199

RESULT 18
US-08-718-547-2
Sequence 2, Application US/08718547
Patent No. 5976794
GENERAL INFORMATION:
APPLICANT: Katz, Aaron E.; Buttyan, Ralph; Raffo, Anthony;
APPLICANT: Olsson, Carl A.
TITLE OF INVENTION: A Method For Molecular Staging Of
TITLE OF INVENTION: Prostate Cancer
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham LLP
STREET: 1185 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/718,547
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 43677-A-PCT/JPW/MJG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 278-0400
TELEFAX: (212) 391-0525
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 237 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-718-547-2

Query Match 80.0%; Score 40; DB 2; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
DB 191 GPLVCNGVL 199

RESULT 19
US-08-768-859A-1
Sequence 1, Application US/08768859A
Patent No. 6013471
GENERAL INFORMATION:
APPLICANT: Tindall, Donald J.
APPLICANT: Young, Charles Y.F.
APPLICANT: Saedi, Mohammed S.
TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
STREET: P.O. Box 2938
CITY: Minneapolis

```
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768, 859A
; FILING DATE: 17-DECEMBER-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-768-859A-1

Query Match 80.0%; Score 40; DB 3; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCNGVL 199

RESULT 20
US-08-768-859A-16
; Sequence 16, Application US/08768859A
; Patent No. 6013471
; GENERAL INFORMATION:
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768, 859A
; FILING DATE: 17-DECEMBER-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
```

```
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-768-859A-16

Query Match 80.0%; Score 40; DB 3; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCNGVL 199

RESULT 21
US-08-768-859A-21
; Sequence 21, Application US/08768859A
; Patent No. 6013471
; GENERAL INFORMATION:
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768, 859A
; FILING DATE: 17-DECEMBER-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-768-859A-21

Query Match 80.0%; Score 40; DB 3; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCNGVL 199

RESULT 22
US-08-767-820A-1
; Sequence 1, Application US/08767820A
; Patent No. 6093796
; GENERAL INFORMATION:
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
```


; APPLICANT: Saeedi, Mohammed S.
; TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/767,820A
; FILING DATE: 17-DECEMBER-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-0331
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-767-820A-1

Query Match 80.0%; Score 40; DB 3; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCGVL 199

RESULT 23
US-08-767-820A-16
; Sequence 16, Application US/08767820A
; Patent No. 6093796
; GENERAL INFORMATION:
; APPLICANT: Tindali, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saeedi, Mohammed S.
; TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/767,820A
; FILING DATE: 17-DECEMBER-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-0331
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-767-820A-16

; REFERENCE/DOCKET NUMBER: 545.002US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-0361
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-767-820A-16

Query Match 80.0%; Score 40; DB 3; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCGVL 199

RESULT 24
US-08-767-820A-21
; Sequence 21, Application US/08767820A
; Patent No. 6093796
; GENERAL INFORMATION:
; APPLICANT: Tindali, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saeedi, Mohammed S.
; TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/767,820A
; FILING DATE: 17-DECEMBER-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-0361
; INFORMATION FOR SEQ ID NO: 21:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-767-820A-21

Query Match 80.0%; Score 40; DB 3; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCGVL 199

RESULT 25
US-08-622-046B-1
; Sequence 1, Application US/08622046B
; Patent No. 6103237
; GENERAL INFORMATION:
; APPLICANT: Saedi, Mohammed S.
; APPLICANT: Mikolajczyk, Stephen D.
; TITLE OF INVENTION: Stable Variant HK2 Polypeptide
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,046B
; FILING DATE: 26-MARCH-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 476.001US1
; TELEPHONE: 612-373-6903
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-622-046B-1

Query Match 80.0%; Score 40; DB 3; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||

Db 191 GPLVCNGVL 199

RESULT 26
US-08-622-046B-7
; Sequence 7, Application US/08622046B
; Patent No. 6103237
; GENERAL INFORMATION:
; APPLICANT: Saedi, Mohammed S.
; APPLICANT: Mikolajczyk, Stephen D.
; TITLE OF INVENTION: Stable Variant HK2 Polypeptide
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,046B

; FILING DATE: 26-MARCH-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 476.001US1
; TELEPHONE: 612-373-6903
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide-
US-08-622-046B-7

Query Match 80.0%; Score 40; DB 3; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||

Db 191 GPLVCNGVL 199

RESULT 27
US-08-622-046B-12
; Sequence 12, Application US/08622046B
; Patent No. 6103237
; GENERAL INFORMATION:
; APPLICANT: Saedi, Mohammed S.
; APPLICANT: Mikolajczyk, Stephen D.
; TITLE OF INVENTION: Stable Variant HK2 Polypeptide
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/622,046B
; FILING DATE: 26-MARCH-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 476.001US1
; TELEPHONE: 612-373-6903
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-622-046B-12

Query Match 80.0%; Score 40; DB 3; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

```
Db 191 GPLVNCVL 199
||||| |
RESULT 28
US-08-944-483-37
; Sequence 37, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 37:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-37
Query Match 80.0%; Score 40; DB 4; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
||||| |
Db 191 GPLVNCVL 199
RESULT 29
US-08-944-483-38
; Sequence 38, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944,483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 38:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-39
Query Match 80.0%; Score 40; DB 4; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
||||| |
Db 191 GPLVNCVL 199
RESULT 30
US-09-100-264-1
; Sequence 1, Application US/09100264
; Patent No. 6235486
; GENERAL INFORMATION:
; APPLICANT: Young, Charles V.F.
; APPLICANT: Tindall, Donald J.
; APPLICANT: Klee, George G.
; TITLE OF INVENTION: Method for detection of breast cancer
; FILE REFERENCE: 545.006US1
; CURRENT APPLICATION NUMBER: US/09/100,264
; CURRENT FILING DATE: 1998-06-19
; EARLIER APPLICATION NUMBER: US 60/050,963
; EARLIER FILING DATE: 1997-06-20
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 237
```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-100-264-1

Query Match 80.0%; Score 40; DB 4; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 0; Indels 2; Gaps 0;

QY 1 GPLVCRGTL 9
||||| |
Db 191 GPLVCGVL 199

RESULT 31

US-09-100-264-3

; Sequence 3, Application US/09100264

; Patent No. 6235486

; GENERAL INFORMATION:

; APPLICANT: Young, Charles Y.F.

; APPLICANT: Tindall, Donald J.

; APPLICANT: Klee, George G.

; TITLE OF INVENTION: Method for detection of breast cancer

; FILE REFERENCE: 545.006US1

; CURRENT APPLICATION NUMBER: US/09/100,264

; CURRENT FILING DATE: 1998-06-19

; EARLIER APPLICATION NUMBER: US 60/050,963

; EARLIER FILING DATE: 1997-06-20

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 3

; LENGTH: 237

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-100-264-3

Query Match

80.0%; Score 40; DB 4; Length 237;

Best Local Similarity 77.8%; Pred. No. 4.9;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||||| |
Db 191 GPLVCGVL 199

RESULT 32

US-09-100-264-12

; Sequence 12, Application US/09100264

; Patent No. 6235486

; GENERAL INFORMATION:

; APPLICANT: Young, Charles Y.F.

; APPLICANT: Tindall, Donald J.

; APPLICANT: Klee, George G.

; TITLE OF INVENTION: Method for detection of breast cancer

; FILE REFERENCE: 545.006US1

; CURRENT APPLICATION NUMBER: US/09/100,264

; CURRENT FILING DATE: 1998-06-19

; EARLIER APPLICATION NUMBER: US 60/050,963

; EARLIER FILING DATE: 1997-06-20

; NUMBER OF SEQ ID NOS: 15

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 12

; LENGTH: 237

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: A variant HK2 polypeptide

US-09-100-264-12

Query Match

80.0%; Score 40; DB 4; Length 237;

Best Local Similarity 77.8%; Pred. No. 4.9;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

Db 191 GPLVCGVL 199
||||| |

RESULT 33

PCT-US94-07329-10

; Sequence 10, Application PC/TUS9407329

; GENERAL INFORMATION:

; APPLICANT: Mayo Foundation for Medical

; APPLICANT: Education and Research

; TITLE OF INVENTION: Antibodies Specific for Human

; TITLE OF INVENTION: Prostate Glandular kallikrein

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Schwegman, Lundberg & Woessner

; STREET: 3500 IDS Center

; STREET: 80 South Eighth Street

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA

; ZIP: 55402

; COMPUTER READABLE FORM: disk

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US94/07329

; FILING DATE:

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: Raasch, Kevin W.

; REGISTRATION NUMBER: 35,651

; REFERENCE/DOCKET NUMBER: 150.62WO

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 612-339-0331

; TELEFAX: 612-339-3061

; INFORMATION FOR SEQ ID NO: 10:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 237 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

; IMMEDIATE SOURCE:

; CLONE: Amino acid sequence of HK2

; CLONE: (deduced from cDNA sequence)

PCT-US94-07329-10

Query Match 80.0%; Score 40; DB 5; Length 237;

Best Local Similarity 77.8%; Pred. No. 4.9;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||||| |
Db 191 GPLVCGVL 199

RESULT 34

PCT-US94-07329-11

; Sequence 11, Application PC/TUS9407329

; GENERAL INFORMATION:

; APPLICANT: Mayo Foundation for Medical

; APPLICANT: Education and Research

; TITLE OF INVENTION: Antibodies Specific for Human

; TITLE OF INVENTION: Prostate Glandular kallikrein

; NUMBER OF SEQUENCES: 11

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Schwegman, Lundberg & Woessner

; STREET: 3500 IDS Center

; STREET: 80 South Eighth Street

; CITY: Minneapolis

; STATE: MN

; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US94/07329
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,561
; REFERENCE/DOCKET NUMBER: 150.62WO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; ORIGINAL SOURCE:
; ORGANISM: Amino acid sequence of PSA (HK3)
; PCT-US94-07329-11

Query Match 80.0%; Score 40; DB 5; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCGVL 199

RESULT 35
PCT-US95-06157-1
; SEQUENCE 16, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saeidi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06157
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,561
; REFERENCE/DOCKET NUMBER: 150.148WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-06157-16

Query Match 80.0%; Score 40; DB 5; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCGVL 199

RESULT 37
US-08-768-859A-8

; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-06157-1

Query Match 80.0%; Score 40; DB 5; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCGVL 199

RESULT 36
PCT-US95-06157-16
; SEQUENCE 16, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saeidi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06157
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,561
; REFERENCE/DOCKET NUMBER: 150.148WO1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 237 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; PCT-US95-06157-16

Query Match 80.0%; Score 40; DB 5; Length 237;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCGVL 199

RESULT 37
US-08-768-859A-8

```
; Sequence 8, Application US/08768859A
; Patent No. 6013471
; GENERAL INFORMATION:
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/768,859A
; FILING DATE: 17-DECEMBER-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US3
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-0331
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-768-859A-8

Query Match 80.0%; Score 40; DB 3; Length 238;
Best Local Similarity 77.8%; Pred. No. 5;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
Db 192 GPLVCGVL 200

RESULT 38
US-08-767-820A-8
; Sequence 8, Application US/08767820A
; Patent No. 6093796
; GENERAL INFORMATION:
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/767,820A
; FILING DATE: 17-DECEMBER-1996
```

```
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US2
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-0331
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-767-820A-8

Query Match 80.0%; Score 40; DB 3; Length 238;
Best Local Similarity 77.8%; Pred. No. 5;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
Db 192 GPLVCGVL 200

RESULT 39
PCT-US95-06157-8
; Sequence 8, Application PC/TUS9506157
; GENERAL INFORMATION:
; APPLICANT: Mayo Foundation for Medical Education
; APPLICANT: and Research
; APPLICANT: Hybritech Incorporated
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saedi, Mohammed S.
; TITLE OF INVENTION: Recombinant HK2 Polypeptide
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner, P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/06157
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Raasch, Kevin W.
; REGISTRATION NUMBER: 35,561
; REFERENCE/DOCKET NUMBER: 150.148WO1
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-0331
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 238 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-06157-8

Query Match 80.0%; Score 40; DB 5; Length 238;
Best Local Similarity 77.8%; Pred. No. 5;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
```

```
Db 192 GPLVCGVL 200
|||||
;
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,831
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/472,228
; FILING DATE: 07-JUN-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSC:388
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-09-146-831-1
;
; Query Match 80.0%; Score 40; DB 4; Length 240;
; Best Local Similarity 77.8%; Pred. No. 5;
; Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
Qy 1 GPLVCRGTL 9
|||||
Db 194 GPLVCGVL 202
|||||
;
; RESULT 42
; PCT-US96-09303-1
; Sequence 1, Application PC/TUS9609303
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES OF PROSTATE
; SPECIFIC ANTIGEN
; NUMBER OF SEQUENCES: 22
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US96/09303
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/472,228
; FILING DATE: 07-JUN-1995
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; PCT-US96-09303-1
;
; Query Match 80.0%; Score 40; DB 5; Length 240;
; Best Local Similarity 77.8%; Pred. No. 5;
; Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
Qy 1 GPLVCRGTL 9
|||||
Db 194 GPLVCGVL 202
|||||
;
; RESULT 41
; US-09-146-831-1
; Sequence 1, Application US/09146831
; Patent No. 6326471
; GENERAL INFORMATION:
; APPLICANT: Kokolus, William J.
; APPLICANT: Fritzsche, Herbert A.
; APPLICANT: Johnston, Dennis A.
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES OF PROSTATE
; SPECIFIC ANTIGEN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
;
; Db 192 GPLVCGVL 200
;
; RESULT 40
; US-08-472-228A-1
; Sequence 1, Application US/08472228A
; Patent No. 5807978
; GENERAL INFORMATION:
; APPLICANT: Kokolus, William J.
; APPLICANT: Fritzsche, Herbert A.
; APPLICANT: Johnston, Dennis A.
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES OF PROSTATE
; SPECIFIC ANTIGEN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
;
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/472,228A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSC:388
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; TELEX: 79-0924
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 240 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-472-228A-1
;
; Query Match 80.0%; Score 40; DB 1; Length 240;
; Best Local Similarity 77.8%; Pred. No. 5;
; Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
;
Qy 1 GPLVCRGTL 9
|||||
Db 194 GPLVCGVL 202
|||||
;
; RESULT 41
; US-09-146-831-1
; Sequence 1, Application US/09146831
; Patent No. 6326471
; GENERAL INFORMATION:
; APPLICANT: Kokolus, William J.
; APPLICANT: Fritzsche, Herbert A.
; APPLICANT: Johnston, Dennis A.
; TITLE OF INVENTION: IMMUNOGENIC PEPTIDES OF PROSTATE
; SPECIFIC ANTIGEN
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
```

Db 194 GPLVCGVL 202
||||| |

RESULT 43

US-08-768-859A-10
; Sequence 10, Application US/08768859A
; Patent No. 6013471
; GENERAL INFORMATION:
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saeedi, Mohammed S.
; TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 17-DECEMBER-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-768-859A-10

Query Match 80.0%; Score 40; DB 3; Length 244;
Best Local Similarity 77.8%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
||||| |

Db 198 GPLVCGVL 206
||||| |

RESULT 44

US-08-767-820A-10
; Sequence 10, Application US/08767820A
; Patent No. 6093796
; GENERAL INFORMATION:
; APPLICANT: Tindall, Donald J.
; APPLICANT: Young, Charles Y.F.
; APPLICANT: Saeedi, Mohammed S.
; TITLE OF INVENTION: DNA Encoding HK2 Variant Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM: disk
US-08-767-820A-10

Query Match 80.0%; Score 40; DB 3; Length 244;
Best Local Similarity 77.8%; Pred. No. 5.1;

; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 17-DECEMBER-1996
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 545.002US2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-767-820A-10

Query Match 80.0%; Score 40; DB 3; Length 244;
Best Local Similarity 77.8%; Pred. No. 5.1;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
||||| |

Db 198 GPLVCGVL 206
||||| |

RESULT 45

US-08-622-046B-5
; Sequence 5, Application US/08622046B
; Patent No. 6103237
; GENERAL INFORMATION:
; APPLICANT: Saeedi, Mohammed S.
; APPLICANT: Mikolajczyk, Stephen D.
; TITLE OF INVENTION: Stable Variant HK2 Polypeptide
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.
; STREET: P.O. Box 2938
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; FILING DATE: 26-MARCH-1996
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Woessner, Warren D.
; REGISTRATION NUMBER: 30,440
; REFERENCE/DOCKET NUMBER: 476.001US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-373-6903
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 244 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-622-046B-5

Query Match 80.0%; Score 40; DB 3; Length 244;
Best Local Similarity 77.8%; Pred. No. 5.1;

Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

Db 198 GPLVCNGVL 206

Search completed: November 6, 2002, 12:06:30
Job time : 9.11111 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 23.7778 Seconds
(without alignments)
42.042 Million cell updates/sec

Title: US-09-905-083-80
Perfect score: 50
Sequence: 1 GPLVCRGTL 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11107396 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_032802:*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	50	100.0	9	22	AAE08285 Human stratum corn
2	50	100.0	225	22	AAE08285 Human stratum corn
3	50	100.0	253	16	AAE08285 Human stratum corn
4	50	100.0	253	17	AAE08285 Human amyloid prec
5	50	100.0	257	21	AAE08285 Human HSCPE. Homo
6	45	90.0	247	22	AAU23217 Novel human enzyme
7	45	90.0	247	22	AAU23217 Novel human enzyme
8	45	90.0	247	22	AAU23217 Novel human enzyme
9	45	90.0	250	21	AAE08285 Human novel secret
10	45	90.0	251	22	AAU16971 Human novel secret
11	45	90.0	296	21	AAE08285 Human KLK-L3 prote

12	44	88.0	237	21	AAE08285	Human KLK-L6 prote
13	44	88.0	251	21	AAE08285	Human KLK-L6 prote
14	42	84.0	115	22	ABG19361	Novel human diagno
15	41	82.0	276	21	AAE08285	Mouse serine prote
16	41	82.0	3201	22	AAE08285	Polyketide synthas
17	41	82.0	3546	22	AAE08285	Polyketide synthas
18	40	80.0	9	20	AAE08285	Immunogenic peptid
19	40	80.0	15	19	AAE08285	Human prostate spe
20	40	80.0	15	19	AAE08285	Human prostate spe
21	40	80.0	20	18	AAE08285	Antigen peptide de
22	40	80.0	125	22	AAE08285	Human colon cancer
23	40	80.0	237	16	AAE08285	Mature kallikrein
24	40	80.0	237	16	AAE08285	Mature kallikrein
25	40	80.0	237	16	AAE08285	Prostate-specific
26	40	80.0	237	19	AAE08285	hK2 variant A217V.
27	40	80.0	237	19	AAE08285	Prostate-specific
28	40	80.0	237	19	AAE08285	Prostate-specific
29	40	80.0	237	19	AAE08285	Mutant human kall
30	40	80.0	237	19	AAE08285	Human prostate spe
31	40	80.0	237	19	AAE08285	Mature prostate sp
32	40	80.0	237	19	AAE08285	Prostate-specific
33	40	80.0	237	20	AAE08285	Mature human kall
34	40	80.0	237	20	AAE08285	Human prostate spe
35	40	80.0	237	20	AAE08285	Variant human kall
36	40	80.0	237	21	AAE08285	Human prostate-spe
37	40	80.0	237	22	AAE08285	Prostate specific
38	40	80.0	237	22	AAE08285	Prostate specific
39	40	80.0	238	16	AAE08285	Mature kallikrein
40	40	80.0	240	18	AAE08285	Human prostate spe
41	40	80.0	244	16	AAE08285	pro-hK2 kallikrein
42	40	80.0	244	19	AAE08285	Prostate-specific
43	40	80.0	244	19	AAE08285	Prostate-specific
44	40	80.0	244	20	AAE08285	Pro human kallikre
45	40	80.0	250	17	AAE08285	Prostate-specific

ALIGNMENTS

RESULT 1
AAE08285
ID AAE08285 standard; peptide; 9 AA.
AC AAE08285;
XX AC
XX AAE08285;
DT 01-NOV-2001 (first entry)
XX Human stratum corneum chymotrypsin enzyme peptide #50 (residues 207-215).
DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antitense therapy; malignant hyperplasia.
XX Homo sapiens.
XX OS
XX WO200159158-A1.
XX 16-AUG-2001.
XX 07-FEB-2001; 2001WO-US03977.
XX 11-FEB-2000; 2000US-0502600.
XX (UYAR-) UNIV ARKANSAS.
XX O'brien TJ;
XX WPI; 2001-514676/56.
XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
PT enzyme -
XX Human novel secret
XX Human KLK-L3 prote
PS Claim 25; Page 113; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
 CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
 CC considered to be an integral part of tumour growth and metastasis, and
 CC therefore, markers indicative of their presence or absence are useful
 CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
 CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
 CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
 CC and other cancers in which SCCE is overexpressed. The present sequence is
 CC human SCCE peptide.

XX Sequence 9 AA;

Query Match 100.0%; Score 50; DB 22; Length 9;
 Best Local Similarity 100.0%; Pred. No. 6.4e+05;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
 Db 1 GPLVCRGTL 9

RESULT 2

AA98502
 ID AAB98502 standard; Protein; 225 AA.

XX AC AAB98502;

XX 03-AUG-2001 (first entry)

XX Human Stratum Corneum Chymotryptic Enzyme, SCCE, catalytic domain.

XX Human: TADG-15; cytostatic; vaccine; ovarian tumour; cancer;
 KW tumour antigen-derived gene 15; serine protease;
 KW Stratum Corneum Chymotryptic Enzyme; SCCE.

XX OS Homo sapiens.

XX PN WO200129056-A1.

XX 26-APR-2001.

XX 20-OCT-2000; 2000WO-US29095.

XX 20-OCT-1999; 99US-0421213.

XX (UYAR-) UNIV ARKANSAS.

XX O'Brien TJ, Tanimoto H;

XX WPI; 2001-381031/40.

XX Novel extracellular serine protease, termed tumor antigen-derived gene
 PT 15 protein overexpressed in carcinomas and DNA encoding it, for
 PT diagnosis, treatment, prevention of cancer, particularly breast,
 PT ovarian cancer

XX Example 10; Fig 1; 130pp; English.

XX The present invention relates to human tumour antigen-derived gene 15
 CC (TADG-15) protein and coding sequence (see AAB23601 and AAB98500).
 CC TADG-15 is an extracellular serine protease. It was found that TADG-15 is
 CC over-expressed in ovarian tumours. TADG-15 protein or its fragments of
 CC 9-20 residues that lack TADG-15 protease activity are useful for
 CC vaccinating an individual against TADG-15, having, suspected of having or
 CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a
 CC diagnostic or therapeutic target in cancer. The present sequence was used
 CC in a sequence homology alignment with the catalytic domain of TADG-15.

XX Sequence 225 AA;

Query Match 100.0%; Score 50; DB 22; Length 225;
 Best Local Similarity 100.0%; Pred. No. 0.78;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 1 GPLVCRGTL 9
 Db 179 GPLVCRGTL 187

RESULT 3

AA67888
 ID AAR67888 standard; Protein; 253 AA.

XX AC AAR67888;

XX 09-AUG-1995 (first entry)

XX Human stratum corneum chymotrophic recombinant enzyme (SCCE).

XX Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
 KW callosities; keratosis pilaris; ichthyoses; eczema.

XX OS Homo sapiens.

XX PN WO9500651-A.

XX 05-JAN-1995.

XX 20-JUN-1994; 94WO-IB00166.

XX 18-JUN-1993; 93DK-0000725.

XX (SYMB-) SYMBICOM AB.

XX Egelrud T, Hansson L;

XX WPI; 1995-052088/07.

XX N-PSDB; AAQ81203.

XX Nucleotide sequences encoding stratum corneum chymotryptic enzyme
 PT - and related vectors, transformed cells and polypeptides, and
 PT useful for treating skin disorders, e.g. acne or psoriasis, and
 PT for identification of specific inhibitors.

XX Disclosure; Page 97; 137pp; English.

XX The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
 CC and skin care products, especially to treat and prevent acne,
 CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
 CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
 CC produced recombinantly following mammal, insect, plant, or
 CC microorganism transformation with plasmid pS507.

XX Sequence 253 AA;

Query Match 100.0%; Score 50; DB 16; Length 253;
 Best Local Similarity 100.0%; Pred. No. 0.87;
 Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
 Db 207 GPLVCRGTL 215

RESULT 4

AAW05383

ID AAW05383 standard; Protein; 253 AA.

XX AC AAW05383;

XX 31-DEC-1996 (first entry)

XX Human amyloid precursor protein protease.

XX Amyloid precursor protein protease; Alzheimer's disease; diagnosis;

KW therapy.
 XX Homo sapiens.
 XX WO9631122-A1.
 XX 10-OCT-1996.
 XX
 XX 02-APR-1996; 96WO-US04294.
 XX
 XX 04-APR-1995; 95US-0416257.
 XX
 XX (ELIL) LILLY & CO ELI.
 XX
 XX Dixon EP, Johnstone EM, Little SP;
 XX WPI; 1996-464694/46.
 XX N-PSDB; AAT39783.
 XX
 XX New isolated human amyloid precursor protein protease - used to
 XX develop prods. for the treatment or diagnosis of associated
 XX conditions, esp. Alzheimer's disease
 XX
 XX Claim 1; Page 44-45; 55pp; English.
 XX
 XX Human amyloid precursor protein protease (AAW05383) is involved in
 XX the processing or clearance of amyloid precursor protein to form
 XX beta-amyloid peptide. Its amino acid sequence was deduced from
 XX a cDNA clone (AAT39783) obtd. from a human lung library. Recombinant
 XX protease can be produced in transformed or transfected prokaryotic
 XX (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
 XX used to develop products for the design and testing of cpds. useful
 XX for treating or preventing conditions associated with beta-amyloid
 XX peptide, esp. Alzheimer's disease.
 XX
 XX SQ Sequence 253 AA;
 XX
 XX Query Match 100.0%; Score 50; DB 17; Length 253;
 XX Best Local Similarity 100.0%; Pred. No. 0.87;
 XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX Qy 1 GPLVCRGRTL 9
 XX | | | | | | | | | |
 XX Db 207 GPLVCRGRTL 215
 XX
 XX RESULT 5
 XX AAB21326
 XX ID AAB21326 standard; Protein: 257 AA.
 XX
 XX AC AAB21326;
 XX
 XX DT 02-FEB-2001 (first entry)
 XX
 XX DE Human HSCEE.
 XX
 XX KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCEE;
 XX KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;
 XX KW serine protease; cytostatic; cancer; prostrate cancer.
 XX
 XX OS Homo sapiens.
 XX
 XX PN WO200053776-A2.
 XX
 XX PD 14-SEP-2000.
 XX
 XX PF 09-MAR-2000; 2000WO-CA00258.
 XX
 XX PR 11-MAR-1999; 99US-0124260.
 XX PR 01-APR-1999; 99US-0127386.
 XX PR 21-JUL-1999; 99US-0144919.
 XX
 XX (MOUN) MOUNT SINAI HOSPITAL.
 XX PA

XX Yousef GM, Diamandis EP;
 XX WPI; 2000-587440/55.
 XX
 XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
 XX protein mediated disorders, especially cancer.
 XX
 XX Example 4; Fig 17; 184pp; English.
 XX
 XX The present sequence is human stratum corneum chymotryptic enzyme
 XX (HSCEE), a member of the kallikrein multi-gene family. Kallikreins and
 XX kallikrein-like proteins are a subgroup of the serine protease enzyme
 XX family. They catalyze the selective cleavage of specific polypeptide
 XX precursors to release peptides with potent biological activity. Nucleic
 XX acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
 XX KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
 XX treatment, monitoring and diagnosis of cancers, especially prostate
 XX cancer. They can also be used to identify a substance that can associate
 XX with or mediate the biological activity of the proteins. Antibodies can
 XX be used to treat conditions mediated by the kallikrein-like proteins.
 XX
 XX SQ Sequence 257 AA;
 XX
 XX Query Match 100.0%; Score 50; DB 21; Length 257;
 XX Best Local Similarity 100.0%; Pred. No. 0.88;
 XX Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 XX
 XX Qy 1 GPLVCRGRTL 9
 XX | | | | | | | | | |
 XX Db 211 GPLVCRGRTL 219
 XX
 XX RESULT 6
 XX AAU23217
 XX ID AAU23217 standard; Protein: 247 AA.
 XX
 XX AC AAU23217;
 XX
 XX DT 18-DEC-2001 (first entry)
 XX
 XX DE Novel human enzyme polypeptide #303.
 XX
 XX KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
 XX KW ligase; hyperproliferative disorder; immunodeficiency disorder;
 XX KW autoimmune disorder; neurological disorder; metabolic disorder;
 XX KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
 XX KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
 XX KW nephrotropic; anticoagulant.
 XX
 XX OS Homo sapiens.
 XX
 XX PN WO200155301-A2.
 XX
 XX PD 02-AUG-2001.
 XX
 XX PF 17-JAN-2001; 2001WO-US01239.
 XX
 XX PR 31-JAN-2000; 2000US-0179065.
 XX PR 04-FEB-2000; 2000US-0180628.
 XX PR 24-FEB-2000; 2000US-0184664.
 XX PR 02-MAR-2000; 2000US-0186350.
 XX PR 16-MAR-2000; 2000US-0189874.
 XX PR 17-MAR-2000; 2000US-0190076.
 XX PR 18-APR-2000; 2000US-0198123.
 XX PR 19-MAY-2000; 2000US-0205515.
 XX PR 07-JUN-2000; 2000US-0209467.
 XX PR 28-JUN-2000; 2000US-0214886.
 XX PR 30-JUN-2000; 2000US-0215135.
 XX PR 07-JUL-2000; 2000US-0216647.
 XX PR 07-JUL-2000; 2000US-0216880.
 XX PR 11-JUL-2000; 2000US-0217487.
 XX PR 11-JUL-2000; 2000US-0217496.

CC (e.g. asthma), cardiovascular disorders (e.g. atherosclerosis),
CC blood-related disorders (e.g. haemophilia), reproductive disorders
CC (e.g. infertility) and infectious disorders (e.g. Influenza). The
CC polynucleotides of the invention can also be used in gene therapy.
CC AAU22915-AAU23814 represent the novel human enzyme polypeptides of the
CC invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 247 AA;

Query Match 90.0%; Score 45; DB 22; Length 247;
Best Local Similarity 88.9%; Pred. No. 5.7;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||
Db 203 GPLVCNGTL 211

RESULT 7
AAU23752
ID AAU23752 standard; Protein; 247 AA.
XX
AC AAU23752;
DT 18-DEC-2001 (first entry)
DE
XX Novel human enzyme polypeptide #838.
KW Human; oxidoreductase enzyme; transferase; hydrolase; lyase; isomerase;
KW ligase; hyperproliferative disorder; immunodeficiency disorder;
KW autoimmune disorder; neurological disorder; metabolic disorder;
KW inflammatory disorder; cardiovascular disorder; reproductive disorder;
KW blood-related disorder; infectious disorder; cytostatic; anti arthritic;
KW nephrotropic; anticoagulant.
XX
OS Homo sapiens.
XX
PN WO200153301-A2.
XX
PD 02-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01239.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228927.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-024617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.

PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 01-SEP-2000; 2000US-0229345.
 PR 03-SEP-2000; 2000US-0229509.
 PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0232080.
 PR 08-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0231968.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 03-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 06-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Rosen CA, Barash SC, Ruben SM;
 XX
 DR WPI; 2001-476222/51.
 DR N-PSDB; AAS26948.
 XX

Novel polypeptides and polynucleotides useful as diagnostic reagents to
 diagnose diseases or disorders associated with aberrant expression or
 activity of polypeptides, for treating blood clotting disorder,
 haemophilia

Claim 11; SEQ ID NO 284; 601pp; English.

The invention relates to isolated nucleic acid molecules and their
 encoded secreted proteins. The nucleic acids and proteins are used to
 prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 rabbits, goats, horses, cats, dogs, chickens or sheep. They
 are also used in diagnosing a pathological condition or susceptibility
 to a pathological condition. Antibodies to the proteins can also
 be used in alleviating symptoms associated with the disorders and in
 diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 immunosorbant assays (ELISA). Disorders which are diagnosed or treated
 include autoimmune diseases e.g. rheumatoid arthritis,
 hyperproliferative disorders e.g. neoplasms of the breast or liver,
 cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.
 Alzheimer's disease, infections caused by bacteria, viruses and fungi
 and ocular disorders e.g. corneal infection, and many other
 disorders listed in the specification. The polypeptides can also
 be used to aid wound healing and epithelial cell proliferation, to
 prevent skin aging due to sunburn, to maintain organs before
 transplantation, for supporting cell culture of primary tissues, to
 regenerate tissues and in chemotaxis. The polypeptides can also be used
 as a food additive or preservative to increase or decrease storage
 capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present

Query Match 90.0%; Score 45; DB 22; Length 247;

Best Local Similarity 88.9%; Pred. No. 5.7;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9

|||||

Db 203 GPLVCGTGL 211

```
RESULT 9
AAB21298
ID AAB21298 standard; Protein; 250 AA.
XX
AC AAB21298;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human KLK-L3 protein #2.
XX
KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6;
KW kallikrein-like protein; serine protease; cytostatic; cancer;
KW prostrate cancer.
XX
OS Homo sapiens.
XX
PN WO200053776-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-CA00258.
XX
PR 11-MAR-1999; 99US-0124260.
PR 01-APR-1999; 99US-0127386.
PR 21-JUL-1999; 99US-0144919.
XX
PA (MOUNT ) MOUNT SINAI HOSPITAL.
XX
PI Yousef GM, Diamandis EP;
DR WPI; 2000-587440/55.
DR N-PSDB; AAA95912.
XX
PT New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
PT protein mediated disorders, especially cancer.
XX
PS Claim 10; Page 155; 184pp; English.
XX
CC The present sequence is kallikrein-like.protein KLK-L3. Kallikreins and
CC kallikrein-like proteins are a subgroup of the serine protease enzyme
CC family. They catalyse the selective cleavage of specific polypeptide
CC precursors to release peptides with potent biological activity. Nucleic
CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
CC treatment, monitoring and diagnosis of cancers, especially prostate
CC cancer. They can also be used to identify a substance that can associate
CC with or mediate the biological activity of the proteins. Antibodies can
CC be used to treat conditions mediated by the kallikrein-like proteins.
XX
SQ Sequence 250 AA;
Query Match 90.0%; Score 45; DB 21; Length 250;
Best Local Similarity 88.9%; Pred. No. 5.8;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
OY 1 GPLVCRGTL 9
| | | | |
Db 206 GPLVCNGTL 214
RESULT 10
AAU16971
ID AAU16971 standard; Protein; 251 AA.
XX
AC AAU16971;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human novel secreted protein, SEQ ID 212.
XX
KW Human; immunosuppressive; antiarthritic; antirheumatic;
```

```
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN WO20015441-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01320.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
```

PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234997.
 PR 25-SEP-2000; 2000US-0234998.
 PR 26-SEP-2000; 2000US-0235484.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0239335.
 PR 13-OCT-2000; 2000US-0239337.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 03-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.

PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251889.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX Rosen CA, Barash SC, Ruben SM;
 PI WPI; 2001-476222/51.
 XX N-PSDB; AAS26876.
 XX Novel polypeptides and polynucleotides useful as diagnostic reagents to
 PT diagnose diseases or disorders associated with aberrant expression or
 PT activity of polypeptides, for treating blood clotting disorder,
 PT haemophilia
 XX
 PS Claim 11; SEQ ID No 212; 601pp; English.
 XX
 CC The invention relates to isolated nucleic acid molecules and their
 CC encoded secreted proteins. The nucleic acids and proteins are used to
 CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,
 CC rabbits, goats, horses, cats, dogs, chickens or sheep. They
 CC are also used in diagnosing a pathological condition or susceptibility
 CC to a pathological condition. Antibodies to the proteins can also
 CC be used in alleviating symptoms associated with the disorders and in
 CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked
 CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated
 CC include autoimmune diseases e.g. rheumatoid arthritis,
 CC hyperproliferative disorders e.g. neoplasms of the breast or liver.
 CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders
 CC e.g. cerebral ischaemia, angioneurosis, nervous system disorders e.g.
 CC Alzheimer's disease, infections caused by bacteria, viruses and fungi
 CC and ocular disorders e.g. corneal infection, and many other
 CC disorders listed in the specification. The polypeptides can also
 CC be used to aid wound healing and epithelial cell proliferation, to
 CC prevent skin aging due to sunburn, to maintain organs before
 CC transplantation, for supporting cell culture of primary tissues, to
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used
 CC as a food additive or preservative to increase or decrease storage
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,
 CC minerals, cofactors and other nutritional components. The present
 CC
 Query Match 90.0%; Score 45; DB 22; Length 251;
 Best Local Similarity 88.9%; Pred. No. 5.8;
 Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 GPLVCRGTL 9
 | | | | |
 Db 207 GPLVCGNGTL 215
 RESULT 11
 AAB21297
 ID AAB21297 standard; Protein; 296 AA.
 XX
 AC AAB21297;
 XX
 DT 02-FEB-2001 (first entry)
 XX
 DE Human KLK-L3 protein #1.
 XX
 KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6;
 KW kallikrein-like protein; serine protease; cytostatic; cancer;
 KW prostrate cancer.
 XX
 OS Homo sapiens.
 XX
 PN WC2000053776-A2.
 XX

PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-CA00258.
XX
XX 11-MAR-1999; 99US-0124260.
PR 01-APR-1999; 99US-0127386.
PR 21-JUL-1999; 99US-0144919.
XX
XX (MOUN) MOUNT SINAI HOSPITAL.
XX
XX Yousef GM, Diamandis EP;
PI WPI; 2000-587440/55.
XX N-PSDB; AAA95912.
DR
XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
PT protein mediated disorders, especially cancer. -
XX
XX Claim 10; Page 155; 184pp; English.
XX
XX The present sequence is kallikrein-like protein KLK-L3. Kallikreins and
CC kallikrein-like proteins are a subgroup of the serine protease enzyme
CC family. They catalyse the selective cleavage of specific polypeptide
CC precursors to release peptides with potent biological activity. Nucleic
CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
CC treatment, monitoring and diagnosis of cancers, especially prostate
CC cancer. They can also be used to identify a substance that can associate
CC with or mediate the biological activity of the proteins. Antibodies can
CC be used to treat conditions mediated by the kallikrein-like proteins.
XX
XX Sequence 296 AA;
SQ
Query Match 90.0%; Score 45; DB 21; Length 296;
Best Local Similarity 88.9%; Pred. No. 6.8;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
Db 252 GPLVCRGTL 260
IIIIII I
RESULT 12
AAB21305
ID AAB21305 standard; Protein; 237 AA.
XX
XX AAB21305;
AC
XX 02-FEB-2001 (first entry)
DT
XX Human KLK-L6 protein #1.
DE
XX Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6;
KW kallikrein-like protein; serine protease; cytostatic; cancer;
KW prostate cancer.
XX
XX Homo sapiens.
OS
XX WO200053776-A2.
PN
XX 14-SEP-2000.
PD
XX 09-MAR-2000; 2000WO-CA00258.
PF
XX 11-MAR-1999; 99US-0124260.
PR 01-APR-1999; 99US-0127386.
PR 21-JUL-1999; 99US-0144919.
XX
XX (MOUN) MOUNT SINAI HOSPITAL.
XX
XX Yousef GM, Diamandis EP;
PI WPI; 2000-587440/55.
XX
XX 14-SEP-2000.
PD
XX 09-MAR-2000; 2000WO-CA00258.
PF
XX 11-MAR-1999; 99US-0124260.
PR 01-APR-1999; 99US-0127386.
PR 21-JUL-1999; 99US-0144919.
XX
XX (MOUN) MOUNT SINAI HOSPITAL.
XX
XX Yousef GM, Diamandis EP;
PI WPI; 2000-587440/55.
XX
DR

DR N-PSDB; AAA95949.
XX
XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
PT protein mediated disorders, especially cancer. -
XX
XX Claim 13; Page 178; 184pp; English.
XX
XX The present sequence is the kallikrein-like protein KLK-L6. Kallikreins
CC and kallikrein-like proteins are a subgroup of the serine protease enzyme
CC family. They catalyse the selective cleavage of specific polypeptide
CC precursors to release peptides with potent biological activity. Nucleic
CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
CC treatment, monitoring and diagnosis of cancers, especially prostate
CC cancer. They can also be used to identify a substance that can associate
CC with or mediate the biological activity of the proteins. Antibodies can
CC be used to treat conditions mediated by the kallikrein-like proteins.
XX
XX Sequence 237 AA;
SQ
Query Match 88.0%; Score 44; DB 21; Length 237;
Best Local Similarity 88.9%; Pred. No. 8.1;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
Db 192 GPLVCRGTL 200
IIIIII I
RESULT 13
AAB21306
ID AAB21306 standard; Protein; 251 AA.
XX
XX AAB21306;
AC
XX 02-FEB-2001 (first entry)
DT
XX Human KLK-L6 protein #2.
DE
XX Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6;
KW kallikrein-like protein; serine protease; cytostatic; cancer;
KW prostate cancer.
XX
XX Homo sapiens.
OS
XX WO200053776-A2.
PN
XX 14-SEP-2000.
PD
XX 09-MAR-2000; 2000WO-CA00258.
PF
XX 11-MAR-1999; 99US-0124260.
PR 01-APR-1999; 99US-0127386.
PR 21-JUL-1999; 99US-0144919.
XX
XX (MOUN) MOUNT SINAI HOSPITAL.
XX
XX Yousef GM, Diamandis EP;
PI WPI; 2000-587440/55.
XX
XX N-PSDB; AAA95949.
DR
XX New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
PT protein mediated disorders, especially cancer. -
XX
XX Claim 13; Page 178; 184pp; English.
XX
XX The present sequence is the kallikrein-like protein KLK-L6. Kallikreins
CC and kallikrein-like proteins are a subgroup of the serine protease enzyme
CC family. They catalyse the selective cleavage of specific polypeptide
CC precursors to release peptides with potent biological activity. Nucleic
CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the
CC treatment, monitoring and diagnosis of cancers, especially prostate
CC cancer. They can also be used to identify a substance that can associate
CC with or mediate the biological activity of the proteins. Antibodies can
CC be used to treat conditions mediated by the kallikrein-like proteins.
XX
XX Sequence 237 AA;
SQ

CC treatment, monitoring and diagnosis of cancers, especially prostate
CC cancer. They can also be used to identify a substance that can associate
CC with or mediate the biological activity of the proteins. Antibodies can
CC be used to treat conditions mediated by the kallikrein-like proteins.

XX Sequence 251 AA;
SQ Query Match 88.0%; Score 44; DB 21; Length 251;
Best Local Similarity 88.9%; Pred. No. 8.5;
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
|||||||
DB 206 GPLVCRGQL 214

RESULT 14
ABG19361
ID ABG19361 standard; Protein; 115 AA.

XX AC ABG19361;

XX DT 18-FEB-2002 (first entry)

XX DE Novel human diagnostic protein #19352.

XX KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX OS Homo sapiens.

XX PN WO200175067-A2.

XX PD 11-OCT-2001.

XX PF 30-MAR-2001; 2001WO-US08631.

XX PR 31-MAR-2000; 2000US-0540217.

XX PR 23-AUG-2000; 2000US-0649167.

XX PA (HYSE-) HYSEQ INC.

XX PI Drmanac RT, Liu C, Tang YT;

XX DR WPI: 2001-639362/73.

XX DR N-PSDB; AAS83548.

XX PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -

XX PS Claim 20; SEQ ID No 49720; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 115 AA;

XX Query Match 84.0%; Score 42; DB 22; Length 115;
Best Local Similarity 77.8%; Pred. No. 8.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
||:|||||
DB 74 GPPICRGTL 82

RESULT 15

XX AAB11713
ID AAB11713 standard; Protein; 276 AA.

XX AC AAB11713;

XX DT 23-OCT-2000 (first entry)

XX DE Mouse serine protease BSSP6 (mBSSP6) SEQ ID NO:4.

XX KW BSSP6; serine protease; human; hBSSP6; mouse; mBSSP6; brain;
KW diagnostic marker; antibody; transgenic animal; Alzheimer's disease;
KW epilepsy; cancer; inflammation; infertility; pancreatitis;
KW prostatic hypertrophy.

XX OS Mus sp.

XX PN WO200031257-A1.

XX PD 02-JUN-2000.

XX PF 19-NOV-1999; 99WO-JP06476.

XX PR 20-NOV-1998; 98JP-0347802.

XX PA (FUSO) FUSO PHARM IND LTD.

XX PI Uemura H, Okui A, Kominami K, Yamaguchi N, Mitsui S;

XX DR WPI: 2000-400067/34.

XX DR N-PSDB; AAA61764.

XX PT Serine protease BSSP6, useful in detecting homologs, mutants and
PT polymorphic variants as markers for diagnosis of Alzheimer's disease,
PT epilepsy, cancer, inflammation, infertility and prostate hypertrophy,
PT using blood or other tissues -

XX PS Claim 3; Page 73-74; 94pp; Japanese.

XX The invention relates to novel serine proteases designated BSSP6
CC (AAB11712-B11714), and to nucleic acids encoding them (AAA61763-AA61765).
CC The invention also relates to vectors and transformants comprising BSSP6
CC nucleic acids; transgenic animals in which the expression level of BSSP6
CC can be varied; and an mBSSP6 knockout mouse. The invention additionally
CC encompasses anti-BSSP6 antibodies and methods of production of such
CC antibodies, methods of BSSP6 detection using the antibodies, and the
CC use of BSSP6 proteins or fragments as diagnostic markers for certain
CC medical conditions. Nucleotides encoding BSSP6 were initially
CC isolated in a human brain cDNA library using degenerate PCR primers
CC (AAA61795-AA61796) based on conserved regions of serine proteases. The
CC BSSP6 serine proteases and nucleotides encoding them are useful in
CC detecting homologues, mutants and polymorphic variants in biological
CC samples (e.g., blood, urine, brain, prostate gland, placenta, testis
CC and spleen) as diagnostic markers for conditions such as Alzheimer's
CC disease, epilepsy, cancer, inflammation, infertility and prostatic
CC hypertrophy. Sequences AAB11712 and AAB11714 represent human BSSP6
CC variants (hBSSP6), and sequence AAB11713 represents murine BSSP6
CC (mBSSP6).

XX

SQ Sequence 276 AA;

Query Match 82.0%; Score 41; DB 21; Length 276;
 Best Local Similarity 77.8%; Pred. No. 29;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 GPLVCRGTL 9
 ||||| I I
 Db 231 GPLVCGSL 239

RESULT 16
 AAB82214
 ID AAB82214 standard; Protein; 3201 AA.
 XX AC AAB82214;
 XX DT 21-JUN-2001 (first entry)
 XX DE Polyketide synthase of the megalomicin biosynthetic pathway.
 XX KW Megalomicin; megAIII gene; polyketide synthase; antibiotic;
 KW megalomicin 6-deoxyerythronolide B synthase 3; motilide;
 KW antiparasitic.
 XX OS Micromonospora megalomicea subsp. nigra.
 XX PN WO200127284-A2.
 XX PD 19-APR-2001.
 XX PF 05-OCT-2000; 2000WO-US27433.
 XX PR 08-OCT-1999; 99US-0158305.
 XX PR 17-MAR-2000; 2000US-0190024.
 XX PA (KOSA-) KOSAN BIOSCIENCES INC.
 XX PI McDaniel R, Volchegursky Y;
 XX DR WPI: 2001-282034/29.
 DR N-PSDB; AAF30757.
 DR P-PSDB; AAB82207, AAB82208, AAB82209, AAB82210, AAB82211, AAB82212,
 DR AAB82213, AAB82214, AAB82215, AAB82216, AAB82217, AAB82218,
 DR AAB82219.
 XX PT Nucleic acids encoding a domain of megalomicin polyketide synthase or a
 PT megalomicin modification enzyme, useful for the production of
 PT megalomicin for use as antibiotics, motilides and antiparasitics. -
 XX Example 1; Page 226-233; 189pp; English.
 XX The present sequence is that of the protein product of the megAIII
 CC gene identified in the megalomicin (meg) gene cluster (see AAF30757)
 CC of Micromonospora megalomicea subsp. nigra (ATCC 27598, NRRL3275).
 CC It is proposed to be a polyketide synthase (PKS), specifically
 CC megalomicin 6-deoxyerythronolide B synthase 3, which shows 83%
 CC similarity to EryAIII. The newly isolated meg gene cluster includes
 CC 3 PKS ORFs encoding enzymes involved in 6-deoxyerythronolide B
 CC (6-DEB) biosynthesis, and other ORFs encoding enzymes involved in
 CC the conversion of 6-DEB to megalomicin or deoxysugar biosynthesis.
 CC The invention provides nucleic acids encoding all or a portion of
 CC the megalomicin PKS and modification enzymes, which can be used to
 CC express recombinant enzymes in host cells for the production of
 CC megalomicin, megalomicin derivatives, and other polyketides,
 CC including hybrid PKS, useful as antibiotics, motilides and
 CC antiparasitics.
 XX SQ Sequence 3201 AA;

Query Match 82.0%; Score 41; DB 22; Length 3201;
 Best Local Similarity 77.8%; Pred. No. 2.7e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 ||||| I I
 Db 1877 GPWVCRGGL 1885

RESULT 17
 AAB82212
 ID AAB82212 standard; Protein; 3546 AA.
 XX AC AAB82212;
 XX DT 21-JUN-2001 (first entry)
 XX DE Polyketide synthase of the megalomicin biosynthetic pathway.
 XX KW Megalomicin; megAI gene; polyketide synthase; antibiotic;
 KW megalomicin 6-deoxyerythronolide B synthase 1; motilide;
 KW antiparasitic.
 XX OS Micromonospora megalomicea subsp. nigra.
 XX PN WO200127284-A2.
 XX PD 19-APR-2001.
 XX PF 05-OCT-2000; 2000WO-US27433.
 XX PR 08-OCT-1999; 99US-0158305.
 XX PR 17-MAR-2000; 2000US-0190024.
 XX PA (KOSA-) KOSAN BIOSCIENCES INC.
 XX PI McDaniel R, Volchegursky Y;
 XX DR WPI: 2001-282034/29.
 DR N-PSDB; AAF30757.
 DR P-PSDB; AAB82207, AAB82208, AAB82209, AAB82210, AAB82211, AAB82212,
 DR AAB82213, AAB82214, AAB82215, AAB82216, AAB82217, AAB82218,
 DR AAB82219.
 XX PT Nucleic acids encoding a domain of megalomicin polyketide synthase or a
 PT megalomicin modification enzyme, useful for the production of
 PT megalomicin for use as antibiotics, motilides and antiparasitics. -
 XX Example 1; Page 211-219; 189pp; English.
 XX The present sequence is that of the protein product of the megAI
 CC gene identified in the megalomicin (meg) gene cluster (see AAF30757)
 CC of Micromonospora megalomicea subsp. nigra (ATCC 27598, NRRL3275).
 CC It is proposed to be a polyketide synthase (PKS), specifically
 CC megalomicin 6-deoxyerythronolide B synthase 1, which shows 81%
 CC similarity to EryAI. The newly isolated meg gene cluster includes
 CC 3 PKS ORFs encoding enzymes involved in 6-deoxyerythronolide B
 CC (6-DEB) biosynthesis, and other ORFs encoding enzymes involved in
 CC the conversion of 6-DEB to megalomicin or deoxysugar biosynthesis.
 CC The invention provides nucleic acids encoding all or a portion of
 CC the megalomicin PKS and modification enzymes, which can be used to
 CC express recombinant enzymes in host cells for the production of
 CC megalomicin, megalomicin derivatives, and other polyketides,
 CC including hybrid PKS, useful as antibiotics, motilides and
 CC antiparasitics.
 XX SQ Sequence 3546 AA;

Query Match 82.0%; Score 41; DB 22; Length 3546;
 Best Local Similarity 77.8%; Pred. No. 3e+02;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Caps 0;

QY 1 GPLVCRGTL 9
 ||||| I I
 Db 2397 GPWVCRGGL 2405

RESULT 18

AAV48056
ID AAY48056 standard; Peptide; 9 AA.

XX
AC AAY48056;

XX
DT 01-DEC-1999 (first entry)

XX
DE Immunogenic peptide having a human leukocyte antigen binding motif #2667.
XX
KW Human leukocyte antigen; binding; immunogenic; glycoprotein; MHC; HLA;
KW Immune response; T cell activation; major histocompatibility complex;
KW cytotoxic T lymphocyte; CTL; tumour rejection; viral infection; cancer;
KW prostate cancer; hepatitis B; hepatitis C; AIDS; renal carcinoma;
KW vaccine; immunisation.

XX
OS Synthetic.

XX
OS Homo sapiens.

XX
PN WO9945954-A1.

XX
PD 16-SEP-1999.

XX
PF 13-MAR-1998; 98WO-US05039.

XX
PR 13-MAR-1998; 98WO-US05039.

XX
PA (EPTM-) EPTIMUNE INC.

XX
PI Sette A, Kubo RT, Sidney J, Celis E, Grey HM, Southwood S;

XX
XX WPI; 1999-551214/46.

XX
PT New immunogenic peptides with HLA binding motif, useful in treatment
PT and diagnosis of cancers and viral diseases -

XX
PS Claim 1; Page 135; 150pp; English.

XX
CC AAY45390 to AAY48214 represent specifically claimed immunogenic peptides
CC having a human major histocompatibility complex (MHC) Class I (also
CC known as human leukocyte antigen (HLA)) binding motif. The immunogenic
CC peptides can bind to a specific HLA allele (i.e. HLA-A subtypes
CC HLA-A2.1, A1, A3.2 or A24.1 or HLA-B or C) and induce a cytotoxic T cell
CC response against the antigen from which the peptide is derived.
CC Cytotoxic T lymphocytes (CTLs) which destroy antigen-bearing cells are
CC normally induced by an antigen in the form of a peptide fragment bound
CC to a HLA molecule, rather than the intact foreign antigen itself, and
CC are particularly important in tumour rejection and in fighting viral
CC infections. The peptides are therefore useful therapeutically to treat
CC or prevent viral infections and cancers in mammals (especially humans)
CC e.g. prostate cancer, hepatitis B and C, AIDS, and renal carcinoma.
CC They can be administered as vaccines to elicit an immune response in
CC individuals susceptible or otherwise at risk of viral infection or
CC cancer, or used to treat chronic or acute conditions. They are also
CC useful diagnostically, and can be used to induce a cytotoxic T cell
CC response, by contacting a cytotoxic T cell with the peptide e.g. to
CC produce CTLs ex vivo for infusion back into a patient. The
CC polynucleotides encoding the immunogenic peptides are also useful
CC therapeutically and for immunisation as above.

XX
SQ Sequence 9 AA;

Query Match 80.0%; Score 40; DB 20; Length 9;
Best Local Similarity 77.8%; Pred. No. 6.4e+05;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

|||||

Db 1 GPLVCGVL 9

RESULT 19

AAW58064

ID AAW58064 standard; peptide; 15 AA.

XX

AAW58063

ID AAW58063 standard; peptide; 15 AA.

XX

AC AAW58063;

XX

DT 11-AUG-1998 (first entry)

XX

DE Human prostate specific antigen peptide SEQ ID NO:81.

XX

KW Human; prostate specific antigen; PSA; epitope; monoclonal antibody;
KW detection; cancer; serine protease.

XX

OS Synthetic.

OS Homo sapiens.

XX
PN WO9810292-A1.

XX
PD 12-MAR-1998.

XX
PF 25-AUG-1997; 97WO-US14909.

XX
PR 06-SEP-1996; 96US-0025404.

XX
PA (CENZ) CENTOCOR INC.

XX
PI Heavner GA;

XX
DR WPI; 1998-193789/17.

XX

PT Monoclonal antibodies specific for prostate specific antigen -
PT useful, e.g. in screening for prostate or breast cancer and
PT especially to distinguish between benign prostatic hyperplasia and
PT prostate cancer

XX
PS Example; Page 64; 84pp; English.

XX
CC The present sequence represents a prostate specific antigen (PSA)
CC peptide. Monoclonal antibodies specific for PSA and hybridomas producing
CC them have been developed. The antibodies: (a) bind to free PSA; (b) are
CC monoclonal antibody 365 binding to amino acids 82-87 of free PSA (LKNRFL)
CC or fragment, and (c) are monoclonal antibodies 10, 11, 16 or 22.2 and
CC 15.2, 156 or 225 binding to amino acids 139-144 (EELFLTP) and 55-60
CC (SLRHPF) respectively of free and bound PSA, or fragments. The antibodies
CC are useful to detect PSA. For detecting free PSA only, an immunoassay
CC comprising a solid support with attached labelled monoclonal antibody
CC specific for free PSA (especially (b)) and a PSA standard can be used.
CC To detect both free and bound PSA, a second solid support with attached
CC (differently labelled) monoclonal antibody binding free and bound PSA
CC (especially selected from (c)) can be used either with, or in place of,
CC the first solid support. The antibodies are useful in cancer screening,
CC especially prostate and breast cancer. By obtaining total and free PSA
CC values, their ratio can be used to separate prostatic cancer (PCA) from
CC benign prostatic hyperplasia (BHP) patients. Measurement of PSA is also
CC useful after radical prostatectomy, to predict disease persistence. The
CC antibodies allow evaluation of PSA free/total ratio, enabling separation
CC of BPH and PCA patients with PSA values 4-10 ng/ml not previously
CC possible by total PSA testing, avoiding biopsies.

XX
SQ Sequence 15 AA;

Query Match 80.0%; Score 40; DB 19; Length 15;
Best Local Similarity 77.8%; Pred. No. 3;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

|||||

Db 5 GPLVCGVL 13

RESULT 20

AAW58064

ID AAW58064 standard; peptide; 15 AA.

XX

```
AC AAW58064;
XX
XX 11-AUG-1998 (first entry)
XX
XX Human prostate specific antigen peptide SEQ ID NO:82.
XX
XX Human; prostate specific antigen; PSA; epitope; monoclonal antibody;
XX detection; cancer; serine protease.
XX
XX Synthetic.
XX Homo sapiens.
XX
XX WO9810292-A1.
XX
XX 12-MAR-1998.
XX
XX 25-AUG-1997; 97NO-US14909.
XX
XX 06-SEP-1996; 96US-0025404.
XX (CENZ ) CENTOCOR INC.
XX
XX Heavner GA;
XX
XX WPI; 1998-193789/17.
XX
XX Monoclonal antibodies specific for prostate specific antigen -
XX useful, e.g. in screening for prostate or breast cancer and
XX especially to distinguish between benign prostatic hyperplasia and
XX prostate cancer
XX
XX Example; Page 64; 84pp; English.
XX
XX The present sequence represents a prostate specific antigen (PSA)
XX peptide. Monoclonal antibodies specific for PSA and hybridomas producing
XX them have been developed. The antibodies: (a) bind to free PSA; (b) are
XX monoclonal antibody 365 binding to amino acids 82-87 of free PSA (LKNRFL)
XX or fragment, and (c) are monoclonal antibodies 10, 11, 16 or 22.2 and
XX 15.2, 156 or 225 binding to amino acids 139-144 (EELFTNP) and 55-60
XX (SLFHPPE) respectively of free and bound PSA, or fragments. The antibodies
XX are useful to detect PSA. For detecting free PSA only, an immunoassay
XX comprising a solid support with attached labelled monoclonal antibody
XX specific for free PSA (especially (b)) and a PSA standard can be used.
XX To detect both free and bound PSA, a second solid support with attached
XX (differently labelled) monoclonal antibody binding free and bound PSA
XX (especially selected from (c)) can be used either with or in place of,
XX the first solid support. The antibodies are useful in cancer screening,
XX especially prostate and breast cancer. By obtaining total and free PSA
XX values, their ratio can be used to separate prostatic cancer (PCA) from
XX benign prostatic hyperplasia (BHP) patients. Measurement of PSA is also
XX useful after radical prostatectomy, to predict disease persistence. The
XX antibodies allow evaluation of PSA free/total ratio, enabling separation
XX of BPH and PCA patients with PSA values 4-10 ng/ml not previously
XX possible by total PSA testing, avoiding biopsies.
XX
XX Sequence 15 AA;
XX
XX Query Match 80.0%; Score 40; DB 19; Length 15;
XX Best Local Similarity 77.8%; Pred. No. 3;
XX Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GPLVCRGTL 9
XX | | | | |
XX Db 2 GPLVCNGVL 10
XX
XX RESULT 21
XX AAW11019
XX ID AAW11019 standard; peptide; 20 AA.
XX
XX AC AAW11019;
XX
XX 27-OCT-1997 (first entry)
XX
XX
```

```
XX
XX Antigen peptide derived from human prostate specific antigen.
XX
XX PSA; kallikrein; prostate cancer; antigenic; polyclonal antisera.
XX
XX Homo sapiens.
XX
XX WO9640754-A1.
XX
XX 19-DEC-1996.
XX
XX 06-JUN-1996; 96WO-US09303.
XX
XX 07-JUN-1995; 95US-0472228.
XX (TEXA ) UNIV TEXAS SYSTEM.
XX
XX Fritsche HA, Johnston DA, Kokolus WJ;
XX WPI; 1997-108633/10.
XX
XX Antigen peptide derived from prostate specific antigen - does not
XX cross react with related kallikreins, for diagnosis of prostate
XX cancer
XX
XX Example 2; Page 42; 74pp; English.
XX
XX The present sequence represents a novel peptide which has 20 contiguous
XX amino acids derived from the 240 residue sequence of the human prostate
XX specific antigen (PSA) (see AAW11023). The preferred peptide has two
XX hydrophobic regions and one hydrophilic region each of about 5 aa in
XX length, arranged as follows: hydrophobic-hydrophilic-hydrophobic.
XX Polyclonal antisera specific for the antigenic peptide may be used in a
XX method for diagnosing prostate cancer in vivo or in vitro. The peptide
XX represents a sequence unique to PSA which does not cross react with
XX certain kallikreins.
XX
XX Sequence 20 AA;
XX
XX Query Match 80.0%; Score 40; DB 18; Length 20;
XX Best Local Similarity 77.8%; Pred. No. 3.9;
XX Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GPLVCRGTL 9
XX | | | | |
XX Db 10 GPLVCNGVL 18
XX
XX RESULT 22
XX AAG75172
XX ID AAG75172 standard; Protein; 125 AA.
XX
XX AC AAG75172;
XX
XX 03-SEP-2001 (first entry)
XX
XX Human colon cancer antigen protein SEQ ID NO:5936.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
XX colorectal carcinoma.
XX
XX Homo sapiens.
XX
XX WO200122920-A2.
XX
XX 05-APR-2001.
XX
XX 28-SEP-2000; 2000WO-US26524.
XX
XX 29-SEP-1999; 99US-0157137.
XX 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
```


XX Ruben SM, Barash SC, Birse CE, Rosen CA;
PI WPI: 2001-235357/24.
DR N-PSDB; AAH34577.
XX
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
PT N-PSDB; AAH34577.
XX
XX Claim 11; Page 7445-7446; 9803pp; English.
XX
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
XX Sequence 125 AA;
Query Match 80.0%; Score 40; DB 22; Length 125;
Best Local Similarity 77.8%; Pred. No. 21;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
DB 113 GPLLCRDTL 121
|||||
RESULT 23
AAR84667
ID AAR84667 standard; Protein; 237 AA.
XX
XX AAR84667;
AC
XX 04-MAR-1996 (first entry)
DT
XX Mature kallikrein hK2.
DE
XX Kallikrein hK2; serine protease; prostate cancer; diagnosis.
KW
XX Homo sapiens.
OS
XX WO9530758-A1.
PN
XX 16-NOV-1995.
PD
XX 09-MAY-1995; 95WO-US06157.
PF
XX 02-MAY-1995; 95US-0427767.
PR
XX 10-MAY-1994; 94US-0241174.
XX
XX (HYBR-) HYBRITECH INC.
PA
XX (MAYO-) MAYO FOUNDATION.
PI
XX Saedi MS, Tindall DJ, Young CYF;
XX WPI: 1995-404123/51.
XX
XX New isolated prostate-specific kallikrein polypeptide(s) - used to
develop prods. for use in assays for such polypeptide(s), partic.
for diagnosis and monitoring of prostate cancer
PT
PT Disclosure; Page 31; 61pp; English.
XX
XX Human kallikrein hK3 (AAR84671) is a prostate-specific kallikrein
CC showing homology to prostate-specific glandular kallikrein hK2
CC (AAR84667). Nonhomologous regions of the kallikreins can be used
CC for prepn. of antibodies specific to hK2.
XX
XX Sequence 237 AA;
Query Match 80.0%; Score 40; DB 16; Length 237;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
DB 113 GPLVCRGTL 199
|||||
RESULT 24
AAR84671
ID AAR84671 standard; Protein; 237 AA.
XX
XX AAR84671;
AC
XX 04-MAR-1996 (first entry)
DT
XX Mature kallikrein hK3.
DE
XX Kallikrein hK3; serine protease; prostate-specific antigen;
KW prostate cancer; diagnosis.
XX
XX Homo sapiens.
OS
XX WO9530758-A1.
PN
XX 16-NOV-1995.
PD
XX 09-MAY-1995; 95WO-US06157.
PF
XX 02-MAY-1995; 95US-0427767.
PR
XX 10-MAY-1994; 94US-0241174.
XX
XX (HYBR-) HYBRITECH INC.
PA
XX (MAYO-) MAYO FOUNDATION.
PI
XX Saedi MS, Tindall DJ, Young CYF;
XX WPI: 1995-404123/51.
XX
XX New isolated prostate-specific kallikrein polypeptide(s) - used to

develop prods. for use in assays for such polypeptide(s), partic.
for diagnosis and monitoring of prostate cancer
PT
XX Disclosure; Page 42; 61pp; English.
PS
XX Human mature hK2 (AAR84667) is a prostate-specific glandular
CC kallikrein useful in the development of prods. used in the diagnosis
CC and monitoring of prostate cancer. Recombinant hK2, contg. an
CC exogenous N-terminal methionine, can be produced in prokaryotic (E.
CC coli) or eukaryotic (insect, mammalian) systems using cDNA (AAT05148)
CC derived from human benign prostatic hyperplasia tissue.
XX
XX Sequence 237 AA;
Query Match 80.0%; Score 40; DB 16; Length 237;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
DB 191 GPLVCRGTL 199
|||||
RESULT 24
AAR84671
ID AAR84671 standard; Protein; 237 AA.
XX
XX AAR84671;
AC
XX 04-MAR-1996 (first entry)
DT
XX Mature kallikrein hK3.
DE
XX Kallikrein hK3; serine protease; prostate-specific antigen;
KW prostate cancer; diagnosis.
XX
XX Homo sapiens.
OS
XX WO9530758-A1.
PN
XX 16-NOV-1995.
PD
XX 09-MAY-1995; 95WO-US06157.
PF
XX 02-MAY-1995; 95US-0427767.
PR
XX 10-MAY-1994; 94US-0241174.
XX
XX (HYBR-) HYBRITECH INC.
PA
XX (MAYO-) MAYO FOUNDATION.
PI
XX Saedi MS, Tindall DJ, Young CYF;
XX WPI: 1995-404123/51.
XX
XX New isolated prostate-specific kallikrein polypeptide(s) - used to
develop prods. for use in assays for such polypeptide(s), partic.
for diagnosis and monitoring of prostate cancer
PT
PT Disclosure; Page 31; 61pp; English.
XX
XX Human kallikrein hK3 (AAR84671) is a prostate-specific kallikrein
CC showing homology to prostate-specific glandular kallikrein hK2
CC (AAR84667). Nonhomologous regions of the kallikreins can be used
CC for prepn. of antibodies specific to hK2.
XX
XX Sequence 237 AA;
Query Match 80.0%; Score 40; DB 16; Length 237;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 GPLVCRGTL 9
DB 113 GPLVCRGTL 199
|||||

```
Db 191 GPLVCGVL 199

RESULT 25
AAR77098
ID AAR77098 standard; Protein; 237 AA.
XX AC
XX AAR77098;
XX DT 16-JAN-1996 (first entry)
XX DE Prostate-specific antigen.
XX KW Prostate-specific antigen; prostate cancer; diagnosis.
XX OS Homo sapiens.
XX PN W09528498-A1.
XX PD 26-OCT-1995.
XX PF 14-APR-1995; 95WO-US04680.
XX PR 15-APR-1994; 94US-0229391.
XX PA (UYCO ) UNIV COLUMBIA NEW YORK.
XX PI Buttyan R, Katz AE, Olsson CA, Raffo A;
XX WPI; 1995-373812/48.
XX DR N-PSDB; AAT04864.
XX PT Method for enhancing prostate-specific antigen detection - provides
XX sensitive means to identify early stages of prostate cancer.
XX PS Disclosure; Page 69-70; 94pp; English.
XX CC A prostate-specific antigen (PSA) is detected by isolating mRNA from
XX a sample, producing cDNA, and subjecting the cDNA to RT-PCR using
XX primers specific for the human PSA coding region (given in AAT04864).
XX SQ Sequence 237 AA;

Query Match 80.0%; Score 40; DB 16; Length 237;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCGVL 199

RESULT 26
AAW83212
ID AAW83212 standard; Protein; 237 AA.
XX AC
XX AAW83212;
XX DT 09-FEB-1999 (first entry)
XX DE hK2 variant A217V.
XX KW Prostate cancer; detection; hK2; hK1; hK3; phK2; tissue kallikrein;
XX pphK2; prostate-specific glandular kallikrein protein; PSA; human;
XX KW prostate-specific antigen.
XX OS Homo sapiens.
XX PN W09846795-A1.
XX PD 22-OCT-1998.
XX PT Detection of human kallikrein 2 RNA - by reverse transcription and
amplification by PCR, for detecting, monitoring and staging of
```

```
PF 09-APR-1998; 98WO-US07027.
XX 11-APR-1997; 97US-0843076.
XX (BAYU ) BAYLOR COLLEGE MEDICINE.
XX PA (MAYO-) MAYO FOUNDATION.
XX PI Slawin KM, Tindall DJ, Young CYF;
XX WPI; 1998-594592/50.
XX DR Detection of human kallikrein 2 RNA - by reverse transcription and
XX amplification by PCR, for detecting, monitoring and staging of
XX PT prostate cancer
XX PS Disclosure; Page 70-71; 90pp; English.
XX CC The present invention describes a diagnostic method for detecting human
XX kallikrein 2 (hK2) DNA. The method comprises: (a) contacting DNA obtained
XX by reverse transcription (RT) of RNA from a human physiological sample
XX which comprises cells suspected of containing hK2 RNA with at least 2
XX oligonucleotides to amplify the DNA by PCR to yield amplified hK2 DNA.
XX CC where the conditions amplify the DNA obtained by RT of RNA from at least
XX one cell containing hK2 in a sample which comprises at least 107 to 109
XX cells; and (b) detecting the presence of the amplified hK2 DNA. The
XX method can be used for detecting, monitoring the progression of and
XX pathologically staging prostate cancer. The present sequence represents
XX a hK2 variant.
XX SQ Sequence 237 AA;

Query Match 80.0%; Score 40; DB 19; Length 237;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
Db 191 GPLVCGVL 199

RESULT 27
AAW83213
ID AAW83213 standard; Protein; 237 AA.
XX AC
XX AAW83213;
XX DT 09-FEB-1999 (first entry)
XX DE Prostate-specific antigen protein hK3.
XX KW Prostate cancer; detection; hK2; hK1; hK3; phK2; tissue kallikrein;
XX pphK2; prostate-specific glandular kallikrein protein; PSA; human;
XX KW prostate-specific antigen.
XX OS Homo sapiens.
XX PN W09846795-A1.
XX PD 22-OCT-1998.
XX PF 09-APR-1998; 98WO-US07027.
XX PR 11-APR-1997; 97US-0843076.
XX PA (BAYU ) BAYLOR COLLEGE MEDICINE.
XX PA (MAYO-) MAYO FOUNDATION.
XX PI Slawin KM, Tindall DJ, Young CYF;
XX WPI; 1998-594592/50.
XX DR Detection of human kallikrein 2 RNA - by reverse transcription and
XX amplification by PCR, for detecting, monitoring and staging of
XX PT
```

The present invention describes a diagnostic method for detecting human kallikrein 2 (hk2) DNA. The method comprises: (a) contacting DNA obtained by reverse transcription (RT) of RNA from a human physiological sample which comprises cells suspected of containing hk2 RNA with at least 2 oligonucleotides to amplify the DNA by PCR to yield amplified hk2 DNA, where the conditions amplify the DNA obtained by RT of RNA from at least one cell containing hk2 in a sample which comprises at least 107 to 109 cells; and (b) detecting the presence of the amplified hk2 DNA. The

Expression of the wild-type hk2 protein (AAW49085) can be detected using hk2 specific antibodies. As hk2 is a marker for prostatic cancer (PC), the method is claimed to be useful for detecting metastatic PC or persistence of PC after radical prostatectomy. Mutant hk2 can be used in assays for detection of wild-type hk2 protein. The method is also claimed to be useful for monitoring the progress of the disease or its

CC treatment and to stage the disease, particularly in subjects being
CC considered for surgery or hormone therapy.

SQ Sequence 237 AA;

Query Match 80.0%; Score 40; DB 19; Length 237;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
||||| |

Db 191 GPLVCGVL 199

RESULT 30

AAW56086

ID AAW56086 standard; Protein; 237 AA.

XX AC AAW56086;

XX 11-AUG-1998 (first entry)

XX Human prostate specific antigen protein.

XX Human; prostate specific antigen; PSA; epitope; monoclonal antibody;
XX detection; cancer; serine protease.

XX Homo sapiens.

XX WO9810292-A1.

XX 12-MAR-1998.

XX 25-AUG-1997; 97WO-US14909.

XX 06-SEP-1996; 96US-0025404.

XX (GENZ) CENTOCOR INC.

XX Heavner GA;

XX WPI; 1998-193789/17.

XX Monoclonal antibodies specific for prostate specific antigen -
XX useful, e.g. in screening for prostate or breast cancer and
XX especially to distinguish between benign prostatic hyperplasia and
XX prostate cancer

PS Disclosure; Page 32-33; 84pp; English.

XX The present sequence represents prostate specific antigen (PSA).
XX Monoclonal antibodies specific for PSA and hybridomas producing them
XX have been developed. The antibodies: (a) bind to free PSA; (b) are
XX monoclonal antibody 365 binding to amino acids 82-87 of free PSA (LKNRFL)
XX or fragment, and (c) are monoclonal antibodies 10, 11, 16 or 22.2 and
XX 15.2, 156 or 225 binding to amino acids 139-144 (EELFLTP) and 55-60
XX (SLFHPPE) respectively of free and bound PSA, or fragments. The antibodies
XX are useful to detect PSA. For detecting free PSA only, an immunoassay
XX comprising a solid support with attached labelled monoclonal antibody
XX specific for free PSA (especially (b)) and a PSA standard can be used.
XX To detect both free and bound PSA, a second solid support with attached
XX (differently labelled) monoclonal antibody binding free and bound PSA
XX (especially selected from (c)) can be used either with, or in place of,
XX the first solid support. The antibodies are useful in cancer screening,
XX especially prostate and breast cancer. By obtaining total and free PSA
XX values, their ratio can be used to separate prostatic cancer (PCA) from
XX benign prostatic hyperplasia (BHP) patients. Measurement of PSA is also
XX useful after radical prostatectomy, to predict disease persistence. The
XX antibodies allow evaluation of PSA free/total ratio, enabling separation
XX of BPH and PCA patients with PSA values 4-10 ng/ml not previously
XX possible by total PSA testing, avoiding biopsies.

SQ Sequence 237 AA;

Query Match 80.0%; Score 40; DB 19; Length 237;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9
||||| |

Db 191 GPLVCGVL 199

RESULT 31

AAW45395

ID AAW45395 standard; Protein; 237 AA.

XX AC AAW45395;

XX 06-JUL-1998 (first entry)

XX Mature prostate-specific glandular kallikrein hk2.

XX Prostate-specific glandular kallikrein; hk2 protein; antigen;
XX prostate carcinoma; prostate cancer; benign prostatic hyperplasia;
XX diagnosis; human.

XX Homo sapiens.

XX WO9802748-A1.

XX 22-JAN-1998.

XX 15-JUL-1997; 97WO-US12322.

XX 15-JUL-1996; 96US-0680868.

XX (HYBR-) HYBRITECH INC.

XX (MAYO-) MAYO FOUNDATION.

XX Grauer L, Klee GG, Mikolajczyk SD, Saedi M, Tindall DJ;

XX Young CYF;

XX WPI; 1998-120378/11.

XX N-PSDB; AAV06602.

XX Diagnostic methods using antibodies which bind prostate antigens -
XX useful for, e.g. monitoring treatment or progression of prostate
XX cancer

PS Example 1; Page 49-50; 100pp; English.

XX This polypeptide comprises human mature prostate-specific glandular
XX kallikrein hk2. It is encoded by cDNA (see AAV06602) derived from
XX human benign prostatic hyperplasia (BPH) tissue RNA. The invention
XX provides a diagnostic method comprising contacting antibodies
XX that specifically bind to pro-hk2 (phk2, see AAW45396) or mature hk2
XX with a sample of physiological fluid from a human. The assay
XX is based on the discovery that phk2 is detected in the
XX supernatant of a prostate carcinoma cell line and that hk2 is
XX present in human physiological fluid from prostate cancer cells.
XX The method is useful for monitoring the treatment and/or
XX progression of prostate cancer, or for the early detection of
XX prostate cancer in males that have BPH or a high grade prostatic
XX neoplasia (HPGN) or whose family members have or had BPH, HPGN or
XX a prostate cancer. Monitoring the presence and/or amount of hk2
XX complexes with plasma proteins may also be important in
XX distinguishing between prostate cancer and BPH.

SQ Sequence 237 AA;

Query Match 80.0%; Score 40; DB 19; Length 237;
Best Local Similarity 77.8%; Pred. No. 37;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 GPLVCRGTL 9


```

XX OS Homo sapiens.
XX PN WO9859073-A1.
XX XX
XX PD 30-DEC-1998.
XX XX
XX PF 19-JUN-1998; 98WO-US12840.
XX PR
XX PR 20-JUN-1997; 97US-0050963.
XX XX
XX PA (KLEE/) KLEE G G.
XX PA (MAYO-) MAYO FOUNDATION.
XX PA (TIND/) TINDALL D J.
XX PA (YOUN/) YOUNG C Y F.
XX PI Klee GG, Tindall DJ, Young CVF;
XX WPI; 1999-105632/09.
XX DR N-PSDB; AAX08947.
XX XX
XX PT Use of human kallikrein 2 - as a marker for developing products for
XX PT the diagnosis, prognosis, monitoring and treatment of breast cancer
XX PS Disclosure; Page 33-34; 67pp; English.
XX XX
XX CC Human kallikrein 2 (hk2) is expressed at elevated levels relative to
XX CC the prostate cancer antigen, prostate specific antigen (PSA) by
XX CC breast cancer cells when stimulated by androgens. Detecting levels
XX CC of hk2 may provide a simple diagnostic tool for detecting or
XX CC determining breast cancer. Detecting hk2 is achieved by producing
XX CC an hk2 DNA from hk2 RNA by reverse transcription. The hk2 DNA
XX CC obtained is then amplified by PCR and detected using probes.
XX CC Determination of breast cancer may also be determined by exposing
XX CC the hk2 polypeptide to an agent which binds to it and then detecting
XX CC directly to the presence of breast cancer in the human individual.
XX CC The methods can be used in the diagnosis, treatment and/or
XX CC monitoring of the progression or remission of breast cancer and/or
XX CC monitoring hk2 levels.
XX XX
XX SQ Sequence 237 AA;
XX
XX Query Match 80.0%; Score 40; DB 20; Length 237;
XX Best Local Similarity 77.8%; Pred. No. 37;
XX Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GPLVCRGTL 9
XX | | | | | |
XX DB 191 GPLVCNGVL 199
XX
XX RESULT 35
XX AAW96190
XX ID AAW96190 standard; Protein; 237 AA.
XX XX
XX AC AAW96190;
XX XX
XX DT 27-APR-1999 (first entry)
XX XX
XX DE Variant human Kallikrein 2 (hk2).
XX XX
XX KW Human kallikrein 2; hk2; breast cancer; diagnosis; detection;
XX KW treatment; monitoring; prostate specific antigen; PSA.
XX XX
XX OS Homo sapiens.
XX XX
XX PN WO9859073-A1.
XX XX
XX PD 30-DEC-1998.
XX XX
XX PF 19-JUN-1998; 98WO-US12840.
XX XX

```

```

PR 20-JUN-1997; 97US-0050963.
XX XX
XX PA (KLEE/) KLEE G G.
XX PA (MAYO-) MAYO FOUNDATION.
XX PA (TIND/) TINDALL D J.
XX PA (YOUN/) YOUNG C Y F.
XX XX
XX PI Klee GG, Tindall DJ, Young CVF;
XX XX
XX WPI; 1999-105632/09.
XX XX
XX PT Use of human kallikrein 2 - as a marker for developing products for
XX PT the diagnosis, prognosis, monitoring and treatment of breast cancer
XX PS Disclosure; Page 42; 67pp; English.
XX XX
XX CC Human kallikrein 2 (hk2) is expressed at elevated levels relative to
XX CC the prostate cancer antigen, prostate specific antigen (PSA) by
XX CC breast cancer cells when stimulated by androgens. Detecting levels
XX CC of hk2 may provide a simple diagnostic tool for detecting or
XX CC determining breast cancer. Detecting hk2 is achieved by producing
XX CC an hk2 DNA from hk2 RNA by reverse transcription. The hk2 DNA
XX CC obtained is then amplified by PCR and detected using probes.
XX CC Determination of breast cancer may also be determined by exposing
XX CC the hk2 polypeptide to an agent which binds to it and then detecting
XX CC directly to the presence of breast cancer in the human individual.
XX CC The methods can be used in the diagnosis, treatment and/or
XX CC monitoring of the progression or remission of breast cancer and/or
XX CC monitoring hk2 levels. This variant of hk2 differs from the wild
XX CC type hk2 (AAW96186) by having an alanine to valine substitution at
XX CC position 217.
XX XX
XX SQ Sequence 237 AA;
XX
XX Query Match 80.0%; Score 40; DB 20; Length 237;
XX Best Local Similarity 77.8%; Pred. No. 37;
XX Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
XX
XX QY 1 GPLVCRGTL 9
XX | | | | | |
XX DB 191 GPLVCNGVL 199
XX
XX RESULT 36
XX AAB11041
XX ID AAB11041 standard; peptide; 237 AA.
XX XX
XX AC AAB11041;
XX XX
XX DT 08-FEB-2001 (first entry)
XX XX
XX DE Human prostate-specific antigen N-terminal fragment #2.
XX XX
XX KW Prostate-specific antigen; PSA; human; detection; prostatic cancer;
XX KW isoform-specific monoclonal antibodies; marker.
XX XX
XX OS Homo sapiens.
XX XX
XX PN EP1043394-A2.
XX XX
XX PD 11-OCT-2000.
XX XX
XX PF 08-APR-2000; 2000EP-0107649.
XX XX
XX PR 10-APR-1999; 99DE-1016242.
XX XX
XX PA (SERA-) SERATEC GES BIOTECHNOLOGIE MBH.
XX XX
XX PI Volk M, Mast W;
XX XX
XX WPI; 2000-674036/66.
XX XX

```

PT Composition containing isoforms of prostate-specific antigen, useful as
 PT a reference standard for analysis and for producing isoform-specific
 XX antibodies -

PS Disclosure; Figure 4; 23pp; German.

XX This invention describes a novel composition (A) having a total
 CC concentration of prostate-specific antigen (PSA). (A) is useful as a
 CC reference in methods for detecting PSA (a marker for prostatic cancer) in
 CC human blood or serum and to generate isoform-specific monoclonal
 CC antibodies (Mab) against PSA, for use in immunological assays. (A)
 CC contains all the isoforms of PSA and can be isolated without significant
 CC formation of artefacts by autoproteolysis. Antibodies specific for
 CC particular PSA isoforms can improve diagnostic specificity.

XX Sequence 237 AA;

Query Match 80.0%; Score 40; DB 21; Length 237;
 Best Local Similarity 77.8%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 ||||| |
 Db 191 GPLVCNGVL 199

RESULT 37

AAB19818
 ID AAB19818 standard; Protein; 237 AA.

XX AC AAB19818;

DT 05-MAR-2001 (first entry)

XX Prostate specific antigen elevated in benign prostatic hyperplasia.

XX Prostate specific antigen; PSA; benign prostatic hyperplasia; BPH;
 KW marker; immunoassay; diagnosis; human.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT Cleavage-site 1 /note= "site of internal peptide bond cleavage"

FT Cleavage-site 145 /note= "site of internal peptide bond cleavage"

FT Cleavage-site 146 /note= "site of internal peptide bond cleavage"

FT Cleavage-site 182 /note= "site of internal peptide bond cleavage"

FT Cleavage-site 182 /note= "site of internal peptide bond cleavage"

XX WO200066718-A1.

XX 09-NOV-2000.

XX 07-APR-2000; 2000WO-US09415.

XX 30-APR-1999; 99US-0303208.

XX (HYBR-) HYBRITECH INC.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Mikolajczyk S, Wang T, Rittenhouse H, Wolfert R, Slawin K;

XX WPI; 2001-015979/02.

XX Novel prostate specific antigen (PSA) for distinguishing benign
 PT prostatic hyperplasia from prostate cancer, comprises a clip at lysine
 PT 182 of the amino acid sequence of a mature form of PSA -

XX Disclosure; Fig 14; 67pp; English.

XX The present sequence represents a mature form of a novel form of

CC prostate specific antigen (PSA). This form of PSA has at least 1
 CC clip at Lys-182 and may additionally have 1 or more clips at Lys-1,
 CC Lys-145 and Lys-146. Preferably, the PSA contains at least 2 clips
 CC at Lys-145 and Lys-182. The novel forms of PSA exist at an
 CC elevated level in patients suspected of having benign prostatic
 CC hyperplasia (BPH) and therefore may be used as a serum marker or
 CC an immunohistological marker to help distinguish BPH from prostate
 CC cancer. Antibodies recognising the novel forms of PSA and
 CC immunoassays that detect and determine the novel forms of PSA are
 CC also provided.

XX Sequence 237 AA;

Query Match 80.0%; Score 40; DB 22; Length 237;
 Best Local Similarity 77.8%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 ||||| |
 Db 191 GPLVCNGVL 199

RESULT 38

AAB19819
 ID AAB19819 standard; Protein; 237 AA.

XX AC AAB19819;

DT 05-MAR-2001 (first entry)

XX Prostate specific antigen specific to benign prostatic hyperplasia.

XX Prostate specific antigen; PSA; BPSA; benign prostatic hyperplasia;
 KW BPH; marker; diagnosis; human.

XX OS Homo sapiens.

XX Key Location/Qualifiers

FT Cleavage-site 1 /note= "site of internal peptide bond cleavage"

FT Cleavage-site 145 /note= "site of internal peptide bond cleavage"

FT Cleavage-site 146 /note= "site of internal peptide bond cleavage"

FT Cleavage-site 182 /note= "site of internal peptide bond cleavage"

FT Cleavage-site 182 /note= "site of internal peptide bond cleavage"

XX WO200067030-A1.

XX 09-NOV-2000.

XX 12-APR-2000; 2000WO-US09789.

XX 30-APR-1999; 99US-0303339.

XX (HYBR-) HYBRITECH INC.

XX (BAYU) BAYLOR COLLEGE MEDICINE.

XX Mikolajczyk S, Wang T, Rittenhouse H, Wolfert R, Slawin K;

XX WPI; 2001-007264/01.

XX Distinguishing prostate cancer from benign prostatic hyperplasia using
 PT different forms of prostate specific antigen contained in a sample
 PT involves mathematically combining amounts of PROPSA and BPSA -

XX Disclosure; Fig 14; 60pp; English.

XX The present sequence represents a mature form of a novel form of
 CC prostate specific antigen (PSA). This form of PSA has at least 1
 CC clip at Lys-182 and may additionally have 1 or more clips at Lys-1,
 CC Lys-145 and Lys-146. BPSA, a form of PSA that comprises at least 1
 CC clip at Lys-182, exists at an elevated level in the transition zone

CC of benign prostatic hyperplasia (BPH) tissue, compared to
 CC peripheral zone cancer and non-cancer prostate tissues. A method
 CC for distinguishing prostate cancer from BPH involves determining
 CC the ratio of proPSA and BPSA in a sample.

SQ Sequence 237 AA;

Query Match 80.0%; Score 40; DB 22; Length 237;
 Best Local Similarity 77.8%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 ||||| |

Db 191 GPLVCGVL 199

RESULT 39

AAR84670
 ID AAR84670 standard; Protein; 238 AA.

XX AC AAR84670;

DT 04-MAR-1996 (first entry)

XX DE Mature kallikrein hK2.

KW Kallikrein hK2; serine protease; prostate cancer; diagnosis.

XX OS Homo sapiens.

FH Key Location/Qualifiers

FT Misc-difference 1
 FT /note= "exogenous N-terminal methionine resulting
 FT from cDNA expression"

XX PN W09530758-A1.

XX PD 16-NOV-1995.

XX PF 09-MAY-1995; 95WO-US06157.

XX PR 02-MAY-1995; 95US-0427767.

XX PR 10-MAY-1994; 94US-0241174.

XX PA (HYBR-) HYBRITTECH INC.

XX PA (MAYO-) MAYO FOUNDATION.

XX PI Saedi MS, Tindall DJ, Young CYF;

XX WPI; 1995-404123/51.

DR N-PSDB; T051489.

XX PT New isolated prostate-specific kallikrein polypeptide(s) - used to
 develop prods. for use in assays for such polypeptide(s), partic.
 for diagnosis and monitoring of prostate cancer

XX PS Disclosure; Page 35-36; 61pp; English.

XX CC Recombinant mature hK2 (AAR84670), contg. an exogenous N-terminal
 methionine, can be produced in prokaryotic (E. coli) or eukaryotic
 (insect, mammalian) systems using cDNA (AAT05148) derived from human
 benign prostatic hyperplasia tissue. hK2 is a prostate-specific
 glandular kallikrein useful in the development of prods. used in
 the diagnosis and monitoring of prostate cancer.

SQ Sequence 238 AA;

Query Match 80.0%; Score 40; DB 16; Length 238;
 Best Local Similarity 77.8%; Pred. No. 37;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9
 ||||| |

Db 192 GPLVCGVL 200

RESULT 40

AAW11023
 ID AAW11023 standard; Protein; 240 AA.

XX AC AAW11023;

DT 27-OCT-1997 (first entry)

XX DE Human prostate specific antigen.

XX PSA; kallikrein; prostate cancer; antigenic; polyclonal antisera.

XX OS Homo sapiens.

XX PN W09640754-A1.

XX PD 19-DEC-1996.

XX PF 06-JUN-1996; 96WO-US09303.

XX PR 07-JUN-1995; 95US-0472228.

XX PA (TEXA) UNIV TEXAS SYSTEM.

XX PI Fritsche HA, Johnston DA, Kokolus WJ;

XX WPI; 1997-108633/10.

XX PT Antigen peptide derived from prostate specific antigen - does not
 cross react with related kallikreins, for diagnosis of prostate
 cancer

XX PS Claim 1; Fig 1; 74pp; English.

XX CC The present sequence represents human prostate specific antigen (PSA)
 which is used to produce novel peptides derived from the 240 residue
 sequence. The preferred peptides have two hydrophobic regions and one
 hydrophilic region each of about 5 aa in length, arranged as follows:
 CC hydrophobic-hydrophilic-hydrophobic. Polyclonal antisera specific for
 the antigenic peptides may be used in a method for diagnosing prostate
 cancer in vivo or in vitro. The peptides represent a sequence unique
 to PSA which does not cross react with certain kallikreins.

SQ Sequence 240 AA;

Query Match 80.0%; Score 40; DB 18; Length 240;
 Best Local Similarity 77.8%; Pred. No. 38;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GPLVCRGTL 9

||||| |

Db 194 GPLVCGVL 202

RESULT 41

AAR84669
 ID AAR84669 standard; Protein; 244 AA.

XX AC AAR84669;

DT 04-MAR-1996 (first entry)

XX DE Pro-hK2 kallikrein.

XX KW Kallikrein hK2; serine protease; prostate cancer; diagnosis.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers
 FT Peptide 1..7

ID AAU29036 standard; Protein; 654 AA.

XX AC AAU29036;

XX DT 18-DEC-2001 (first entry)

XX DE Human PRO polypeptide sequence #13.

XX KW PRO polypeptide; mammal; tumour; cancer; human; cattle; horse; sheep; dog; cat; pig; goat; rabbit; tumour necrosis factor alpha; TNF-alpha; blood; chondrocyte cell; cell proliferation; cell differentiation; colon; adrenal; lung; breast; prostate; cervix; liver; genetic disorder.

XX OS Homo sapiens.

XX PN WC20016848-A2.

XX PD 20-SEP-2001.

XX PF 28-FEB-2001; 2001WO-US06520.

XX PR 01-MAR-2000; 2000WO-US05601.

XX PR 02-MAR-2000; 2000WO-US05841.

XX PR 03-MAR-2000; 2000US-187202P.

XX PR 06-MAR-2000; 2000US-186968P.

XX PR 14-MAR-2000; 2000US-189320P.

XX PR 14-MAR-2000; 2000US-189328P.

XX PR 15-MAR-2000; 2000WO-US06884.

XX PR 21-MAR-2000; 2000US-191007P.

XX PR 21-MAR-2000; 2000US-191048P.

XX PR 21-MAR-2000; 2000US-191314P.

XX PR 28-MAR-2000; 2000US-192655P.

XX PR 29-MAR-2000; 2000US-193032P.

XX PR 29-MAR-2000; 2000US-193053P.

XX PR 30-MAR-2000; 2000WO-US08439.

XX PR 04-APR-2000; 2000US-194449P.

XX PR 04-APR-2000; 2000US-194647P.

XX PR 11-APR-2000; 2000US-195975P.

XX PR 11-APR-2000; 2000US-196000P.

XX PR 11-APR-2000; 2000US-196187P.

XX PR 11-APR-2000; 2000US-196690P.

XX PR 18-APR-2000; 2000US-196820P.

XX PR 18-APR-2000; 2000US-198121P.

XX PR 18-APR-2000; 2000US-198585P.

XX PR 25-APR-2000; 2000US-199397P.

XX PR 25-APR-2000; 2000US-199550P.

XX PR 03-MAY-2000; 2000US-199654P.

XX PR 17-MAY-2000; 2000US-201516P.

XX PR 22-MAY-2000; 2000WO-US14042.

XX PR 30-MAY-2000; 2000WO-US14941.

XX PR 02-JUN-2000; 2000WO-US15264.

XX PR 05-JUN-2000; 2000US-209832P.

XX PR 28-JUL-2000; 2000WO-US20710.

XX PR 22-AUG-2000; 2000US-0644848.

XX PR 24-AUG-2000; 2000WO-US23328.

XX PR 08-NOV-2000; 2000WO-US30952.

XX PR 01-DEC-2000; 2000WO-US32678.

XX PR 20-DEC-2000; 2000WO-US34956.

XX (GETH) GENENTECH INC.

XX Baker KP, Chen J, Desnoyers L, Goddard A, Godowski PJ, Gurney AL;

XX Pan J, Smith V, Watanabe CK, Wood WI, Zhang Z;

XX WPI; 2001-602746/68.

XX N-PSDB; AAS45937.

XX Novel nucleic acids encoding PRO polypeptides, used to diagnose the

XX presence of tumours, such as prostate and breast tumours, in mammals and

XX to screen for modulators of the compounds -

XX

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

Claim 11; Fig 26; 774pp; English.

Sequences AAU29024-AAU29328 represent PRO polypeptides of the invention. The PRO polypeptides and their associated nucleic acids can be used to detect the presence of a tumour in a mammal by comparing the level of expression of a PRO polypeptide in a test sample of cells from the animal and a control sample of normal cells, whereby a higher level of expression in the test sample indicates the presence of a tumour in the mammal. Mammals include dogs, cats, cattle, horses, sheep, pigs, goats and rabbits but are preferably human. The polypeptides can be used to stimulate tumour necrosis factor (TNF) alpha release from human blood, when contacted with it. A specific polypeptide can be used to stimulate the proliferation or differentiation of chondrocyte cells. The PRO proteins can be used to determine the presence of tumours and also susceptibility to tumour development, particularly adrenal, lung, colon, breast, prostate, rectal, cervical, or liver tumours, in mammalian subjects. The oligonucleotide probes specific for the PRO nucleic acids can be used for genetic analysis of individuals with genetic disorders.

Sequence 654 AA;

Query Match 78.0%; Score 32; DB 22; Length 654;

Best Local Similarity 100.0%; Pred. No. 2.7e+02;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 RSLLLPL 9

Db 10 RSLLLPL 16

RESULT 44

AAE06593

ID AAE06593 standard; Protein; 654 AA.

XX AC AAE06593;

XX DT 25-SEP-2001 (first entry)

XX DE Human protein having hydrophobic domain, HP03934.

XX KW Human; hydrophobic domain; gene therapy; nutritional supplement; cell proliferation; immunomodulatory; autoimmune disorder; antimicrobial; multiple sclerosis; rheumatoid arthritis; insulin-dependent diabetes; haematopoiesis; tissue growth activity; Parkinson's disease; cytostatic; Huntington's disease; Alzheimer's disease; chemotactic; chemokinetic; haemostatic; thrombolytic; tumour growth inhibitor; anabolic; contraceptive; antiinfertility; antiinflammatory.

XX OS Homo sapiens.

XX FH Key Location/Qualifiers

XX FT Peptide 1..28 /label= Signal_peptide

XX FT Protein 29..654 /note= "Mature human protein with hydrophobic domain"

XX PN WC200149728-A2.

XX PD 12-JUL-2001.

XX PF 28-DEC-2000; 2000WO-JP09359.

XX PR 06-JAN-2000; 2000JP-0000585.

XX PR 06-JAN-2000; 2000JP-0000588.

XX PR 11-JAN-2000; 2000JP-0002299.

XX PR 03-FEB-2000; 2000JP-0026862.

XX PR 03-MAR-2000; 2000JP-0058367.

XX (PROT-) PROTEGENE INC.

XX PA (SAGA) SAGAMI CHEM RES CENT.

XX PI Kato S, Kimura T;

XX

DR WPI; 2001-418355/44.
XX N-PSDB; AAD12588.
XX Human proteins with hydrophobic domains and the nucleic acids encoding
PT them, useful for preventing diagnosing and treating e.g. cancer,
PT Alzheimer's and inflammation -
XX Claim 1; Page 100-101; 563pp; English.
XX The present sequence is human protein with hydrophobic domain,
CC HP03934. The polynucleotide and polypeptide of the invention
CC may be used in the prevention, diagnosis and treatment of diseases
CC associated with inappropriate polypeptide expression. The polynucleotides
CC may be used to produce the polypeptide, by inserting the nucleic acids
CC into a host cell and culturing the cell to express the protein. The
CC polynucleotides and its complementary sequences may also be used as DNA
CC probes in diagnostic assays and also used in gene therapy. The
CC polypeptides may also be used as antigens in the production of antibodies
CC and in assays to identify modulators of polypeptide expression and
CC activity. The polypeptides and nucleic acids may be used as nutritional
CC supplements, to modulate cytokine and cell proliferation activity, to
CC modulate immune stimulation or suppression (e.g. for the treatment of
CC microbial infections and autoimmune disorders such as multiple sclerosis,
CC rheumatoid arthritis and insulin-dependent diabetes), to modulate
CC haematopoiesis, to modulate tissue growth activity (e.g. for the
CC treatment of Parkinson's disease, Huntington's disease and Alzheimer's
CC disease), to modulate actin and huntingtin activity (e.g. for controlling
CC fertility), to modulate chemotactic and chemokinetic activity, to
CC modulate haemostatic and thrombolytic activity, to modulate receptor
CC ligand activity, to modulate inflammation and to inhibit tumour growth.
XX
SQ Sequence 654 AA;
Query Match 78.0%; Score 32; DB 22; Length 654;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 RSLLLPL 9
Db 10 RSLLLPL 16
RESULT 45
AAB80243
ID AAB80243 standard; Protein; 654 AA.
XX
AC AAB80243;
XX
DT 24-APR-2001 (first entry)
XX
DE Human PRO262 protein.
XX
KW Human; PRO; dermatological; antipsoriatic; cytostatic; antiinflammatory;
KW antiparkinsonian nootropic; neuroprotective; vulnary; cardiant;
KW antiangiogenic; vasotropic; antiasthmatic; antirheumatic; cancer;
KW antiarthritic; antifertility; antidiabetic; antiviral; diabetes;
KW ophthalmological; gene therapy; skin disease; gastrointestinal disorder;
KW ischaemia; inflammation.
XX
OS Homo sapiens.
XX
PN WO200104311-A1.
XX
PD 18-JAN-2001.
XX
PF 22-FEB-2000; 2000WO-US04414.
XX
PR 07-JUL-1999; 99US-0143048.
PR 26-JUL-1999; 99US-0145698.
PR 28-JUL-1999; 99US-0146222.
PR 08-SEP-1999; 99WO-US20594.
PR 13-SEP-1999; 99WO-US20944.
PR 15-SEP-1999; 99WO-US21090.

PR 15-SEP-1999; 99WO-US21547.
PR 05-OCT-1999; 99WO-US23089.
PR 29-NOV-1999; 99WO-US28214.
PR 30-NOV-1999; 99WO-US28313.
PR 16-DEC-1999; 99WO-US30095.
PR 20-DEC-1999; 99WO-US30911.
PR 20-DEC-1999; 99WO-US30999.
PR 05-JAN-2000; 99WO-US00219.
XX
XX (GETH) GENENTECH INC.
XX
PI Ashkenazi AJ, Botstein D, Desnoyers L, Eaton DL, Ferrara N;
PI Filvaroff E, Fong S, Gao W, Gerber H, Gerritsen ME, Goddard A;
PI Godowski PJ, Grimaldi CJ, Gurney AL, Hillan KJ, Kljavin IJ;
PI Mather JP, Pan J, Paoni NF, Roy MA, Stewart TA, Tumas D;
PI Williams PM, Wood WI;
DR WPI; 2001-081051/09.
DR N-PSDB; AAF72404.
XX
XX Sixty one nucleic acids encoding PRO polypeptides which are useful in
PT the treatment of skin diseases (e.g. psoriasis), cancers (e.g. lung
PT squamous cell carcinoma) and neurodegenerative diseases (e.g.
PT Alzheimer's disease) -
XX
PS Claim 1; Fig 66; 393pp; English.
XX
CC The present sequence is one of sixty one novel secreted and
CC transmembrane PRO polypeptides. The PRO polypeptides are
CC useful for treating skin diseases (e.g. psoriasis), cancers (e.g. lung
CC squamous cell carcinoma), gastrointestinal disorders (e.g.
CC enterocolitis), neurodegenerative diseases (e.g. Alzheimer's disease,
CC Parkinson's disease), wound repair, cardiovascular disorders (e.g.
CC endometrial bleeding angiogenesis, ischaemias such as coronary
CC ischaemia, atherosclerosis), inflammatory disorders (e.g. asthma,
CC rheumatoid arthritis, multiple sclerosis), infertility, AIDS and
CC diabetes and retinal disorders such as retinitis pigmentosum.
CC The PRO nucleic acids have applications in molecular biology, including
CC use as hybridization probes, and in chromosome and gene mapping.
XX
SQ Sequence 654 AA;
Query Match 78.0%; Score 32; DB 22; Length 654;
Best Local Similarity 100.0%; Pred. No. 2.7e+02;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 3 RSLLLPL 9
Db 10 RSLLLPL 16

Search completed: November 6, 2002, 12:05:16
Job time : 26.7778 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 18.4444 Seconds
(without alignments)
84.413 Million cell updates/sec

Title: US-09-905-083-99
Perfect score: 42
Sequence: 1 QRIKASKSF 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	36	85.7	234	11 Q9R048	Q9R048 mus musculu
2	36	85.7	249	11 Q9LVE3	Q9LVE3 mus musculu
3	34	81.0	107	16 Q9CJU4	Q9CJU4 pasteurella
4	33	78.6	427	4 Q96BZ5	Q96BZ5 homo sapien
5	32	76.2	274	5 Q9U023	Q9U023 leishmania
6	32	76.2	895	17 Q9YAU1	Q9YAU1 aeropyrum p
7	31	73.8	248	2 Q937X0	Q937X0 sphingomonas
8	31	73.8	254	17 Q975X7	Q975X7 sulfolobus
9	31	73.8	302	3 Q9UTE6	Q9UTE6 schizosacch
10	31	73.8	362	16 Q9PHQ2	Q9PHQ2 campylobact
11	30	71.4	321	16 Q84847	Q84847 chlamydia t
12	30	71.4	331	16 Q97JP2	Q97JP2 clostridium
13	30	71.4	427	10 Q9M035	Q9M035 arabidopsis
14	30	71.4	478	16 Q926K2	Q926K2 listeria in
15	30	71.4	583	10 Q9FM26	Q9FM26 arabidopsis
16	30	71.4	583	10 Q93ZP5	Q93ZP5 arabidopsis

17	30	71.4	726	4	Q15152	Q15152 homo sapien
18	30	71.4	727	6	Q28161	Q28161 bos taurus
19	30	71.4	728	11	P97350	P97350 mus musculu
20	30	71.4	826	5	Q968Z5	Q968Z5 caenorhabdi
21	30	71.4	875	5	Q18372	Q18372 caenorhabdi
22	30	71.4	985	16	Q9PFI8	Q9PFI8 xyella fas
23	30	71.4	1102	10	Q9ZVX1	Q9ZVX1 arabidopsis
24	29	69.0	140	12	P89914	P89914 tobacco rat
25	29	69.0	149	5	O00897	O00897 dictyosteli
26	29	69.0	172	16	Q9PMA5	Q9PMA5 campylobact
27	29	69.0	175	10	Q94JH8	Q94JH8 oryza sativ
28	29	69.0	182	5	Q9V6Z7	Q9V6Z7 drosophila
29	29	69.0	201	12	Q91B14	Q91B14 spodoptera
30	29	69.0	202	6	Q95KC4	Q95KC4 macaca fasc
31	29	69.0	236	10	Q94A31	Q94A31 arabidopsis
32	29	69.0	244	13	O42159	O42159 petromyzon
33	29	69.0	245	13	O42160	O42160 petromyzon
34	29	69.0	268	17	Q9HNY9	Q9HNY9 halobacteri
35	29	69.0	308	5	Q9VJG8	Q9VJG8 drosophila
36	29	69.0	325	5	Q9GRJ8	Q9GRJ8 leishmania
37	29	69.0	368	3	O07224	O07224 saccharomyc
38	29	69.0	368	5	O62258	O62258 caenorhabdi
39	29	69.0	439	16	Q9A1I9	Q9A1I9 streptococc
40	29	69.0	463	16	Q97E56	Q97E56 clostridium
41	29	69.0	468	2	Q9L0P9	Q9L0P9 streptomyce
42	29	69.0	526	17	O58616	O58616 pyrococcus
43	29	69.0	563	16	Q9PPZ1	Q9PPZ1 uraeoplasma
44	29	69.0	584	10	Q9LZ50	Q9LZ50 arabidopsis
45	29	69.0	669	5	Q9V6Z8	Q9V6Z8 drosophila

ALIGNMENTS

RESULT 1

Q9R048 ID Q9R048 PRELIMINARY; PRT; 234 AA.
AC Q9R048;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE STRATUM CORNEUM CHYMOTRYPTIC ENZYME (FRAGMENT).
GN K1K7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL;
RX MEDLINE=99399282; PubMed=10469296;
RA Backman A., Stranden P., Brattsand M., Hansson L., Egelrud T.;
RT "Molecular cloning and tissue expression of the murine analog to human
J. Invest. Dermatol. 113:152-155 (1999).
CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
CC EMBL; AF124299; AAF01139.1; -.
DR HSSP; P00757; ISGF.
DR MEROPS; S01.300; -.
DR MGD; MGI:1346336; K1K7.
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; Trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYP.
DR SMART; SM00020; TRYP_SPC; 1.
DR PROSITE; PS00240; TRYPIN_DOM; 1.
DR PROSITE; PS00134; TRYPIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPIN_SER; 1.
KW Hydrolase; Serine protease.
FT NON_TER 234 234
SQ SEQUENCE 234 AA; 25220 MW; BEBE57D6C325B85F CRC64;

Query Match 85.7%; Score 36; DB 11; Length 234;
Best Local Similarity 77.8%; Pred. No. 4.6;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
I:::|||||
Db 87 QKIKATKSF 95

RESULT 2
Q91VE3
ID Q91VE3 PRELIMINARY; PRT; 249 AA.
AC Q91VE3;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE THYMOPIN (STRATUM CORNEUM CHYMOTRYPTIC ENZYME).
GN KLK7.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=THYMUS;
RA Yamaguchi N.;
RT "A novel cDNA cloning of mouse serine protease, Thymopain, i.";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RA Hansson L., Backman A., Ny A., Edlund M., Edholm E., Tornell J.,
RA Wallbrandt P., Egelrud T.;
RT "Epidermal overexpression of stratum corneum chymotryptic enzyme in
RT mice: a model for chronic itchy dermatitis.";
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB008371; BAB55604.1; -;
DR EMBL; AF339930; AAK69652.1; -;
SQ SEQUENCE 249 AA; 27257 MW; 0D4E380F12D14F87 CRC64;

Query Match 85.7%; Score 36; DB 11; Length 249;
Best Local Similarity 77.8%; Pred. No. 4.9;
Matches 7; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
I:::|||||
Db 87 QKIKATKSF 95

RESULT 3
Q9CJU4
ID Q9CJU4 PRELIMINARY; PRT; 107 AA.
AC Q9CJU4;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN PM1897.
GN PM1897.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Pm70.";
RL proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR EMBL; AE006227; AAK03981.1; -;
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 107 AA; 12214 MW; 920C61E14A10F405 CRC64;

Query Match 81.0%; Score 34; DB 16; Length 107;
Best Local Similarity 77.8%; Pred. No. 5.9;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
I:::|||||
Db 27 QRIEAGKSF 35

RESULT 4
Q96BZ5
ID Q96BZ5 PRELIMINARY; PRT; 427 AA.
AC Q96BZ5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 48.5 KDA PROTEIN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON ADENOCARCINOMA;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014992; AAH14992.1; -;
KW Hypothetical protein.
SQ SEQUENCE 427 AA; 48542 MW; 68EBE7AF956BF77 CRC64;

Query Match 78.6%; Score 33; DB 4; Length 427;
Best Local Similarity 66.7%; Pred. No. 40;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
I:::|||||
Db 355 QKLEASKSF 363

RESULT 5
Q900Z3
ID Q900Z3 PRELIMINARY; PRT; 274 AA.
AC Q900Z3;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 29.9 KDA PROTEIN.
GN L5883.04.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Murphy L., Harris D., Ivens A.C., Lawson D., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
DR EMBL; AL117384; CAB55615.1; -;
DR InterPro; IPR000330; SNF2_N.
DR Pfam; PF00176; SNF2_N; 1.
KW Hypothetical protein.
SQ SEQUENCE 274 AA; 29871 MW; E37022F9F935FA0C CRC64;

Query Match 76.2%; Score 32; DB 5; Length 274;
Best Local Similarity 66.7%; Pred. No. 42;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
||||:|
Db 175 QRIKCARSF 183

RESULT 6

Q9YAU1 PRELIMINARY; PRT; 895 AA.
AC Q9YAU1:
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE 895AA LONG HYPOTHETICAL DNA-DIRECTED RNA POLYMERASE SUBUNIT A'.
GN APEI853.
OS Aeropyrum pernix.
OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
OC Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K1;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayasi Y., Hino Y., Horikawa H., Yamazaki S., Haikawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kushiida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix K1.";
RL DNA Res. 6:83-101(1999).
DR EMBL: AP000062; BAA80857.1; -.
DR InterPro: IPR000722; RNA_pol_A.
DR Fram: FR00623; RNA_pol_A: 1.
KW DNA-directed RNA polymerase; Complete proteome.
SQ SEQUENCE 895 AA; 101740 MW; 1E01BC06B4BD0AAE CRC64;

Query Match 76.2%; Score 32; DB 17; Length 895;
Best Local Similarity 75.0%; Pred. NO. 1.4e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
||:|
Db 885 QRVKASKA 892

RESULT 7

Q937X0 PRELIMINARY; PRT; 248 AA.
AC Q937X0:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TETRACHLOROHYDROQUINONE REDUCTIVE DEHALOGENASE.
GN PCPC.
OS Sphingomonas sp. UG30.
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Sphingomonas.
OX NCBI_TaxID=76070;
RN [1]
RP SEQUENCE FROM N.A.
RA Habash M.B., Beaudette L.A., Cassidy M.B., Lee H., Trevors J.T.;
RT "Cloning and Characterization of a Tetrachlorohydroquinone Reductive
RT Dehalogenase from Sphingomonas sp. UG30.";
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY057901; AAL24440.1; -.
SQ SEQUENCE 248 AA; 28247 MW; 850B009C6013426C CRC64;

Query Match 73.8%; Score 31; DB 2; Length 248;
Best Local Similarity 66.7%; Pred. NO. 64;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
||:|:|
Db 227 QRMKARRSF 235

RESULT 8

Q975X7 PRELIMINARY; PRT; 254 AA.
AC Q975X7:
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL PROTEIN ST0301.
GN ST0301.
OS Sulfolobus tokodaii.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=111955;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JCM 10545 / 7;
RX PubMed=11572479;
RA Kawarabayasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,
RA Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S.,
RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,
RA Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Kushiida N., Oguchi A.,
RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,
RA Oshima T., Kikuchi H.;
RT "Complete genome sequence of an aerobic thermoacidophilic
RT crenarchaeon, Sulfolobus tokodaii strain 7.";
RL DNA Res. 8:123-140(2001).
DR EMBL: AP000982; BAB65271.1; -.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 254 AA; 28925 MW; 83208D0BD69A6929 CRC64;

Query Match 73.8%; Score 31; DB 17; Length 254;
Best Local Similarity 75.0%; Pred. NO. 66;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 RIKASKSF 9
|||:|:|
Db 175 RIKATKNF 182

RESULT 9

Q9UTE6 PRELIMINARY; PRT; 302 AA.
AC Q9UTE6:
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 35.6 KDA PROTEIN C222.06 IN CHROMOSOME I.
GN SPAC222.06.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomyces.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972;
RA Hamlin N., Churcher C.M., McDougall R.C., Rajandream M.A.,
RA Barrell B.G.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: STRONG, TO YEAST MAK16.
DR EMBL: AL132798; CAB60698.1; -.
KW Hypothetical protein.
FT DOMAIN 199 210 POLY-GLU.
FT DOMAIN 240 245 POLY-GLU.
SQ SEQUENCE 302 AA; 35626 MW; 5C08899FF29C0E67 CRC64;

Query Match 73.8%; Score 31; DB 3; Length 302;
Best Local Similarity 66.7%; Pred. NO. 78;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 QRIKASKSF 9
 |||| ||:::
 Db 78 QRIKLSKNY 86

RESULT 10

ID Q9PHQ2 PRELIMINARY; PRT; 362 AA.
 AC Q9PHQ2;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE PHOSPHATE TRANSPORT SYSTEM PERMEASE PROTEIN.
 GN PSTA OR C30615.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NCTC 11168;
 RX MEDLINE=20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Ketley J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jagels K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni
 RT reveals hypervariable sequences.";
 RL Nature 403:665-668(2000).
 DR EMBL; AL139075; CAB75251.1; -;
 DR InterPro; IPR000515; BPD.transp.
 DR Pfam; PF00528; BPD_transp; 1.
 DR Complete proteome.
 KW Campylobacter.
 SQ SEQUENCE 362 AA; 39726 MW; 988C797D104F8CCF CRC64;

Query Match 73.8%; Score 31; DB 16; Length 362;
 Best Local Similarity 77.8%; Pred. No. 94;
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9

:| |||||

Db 7 KRIKASKSF 15

RESULT 11

ID O84847 PRELIMINARY; PRT; 321 AA.
 AC O84847;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PP-LOOP SUPERFAMILY ATPASE.
 GN MESJ OR CT840.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UW-3/CX;
 RX MEDLINE=99000809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 DR EMBL; AE001357; AAC68437.1; -;
 DR Complete proteome.
 KW Chlamydia.
 SQ SEQUENCE 321 AA; 37332 MW; 0786CF9211C26CEE CRC64;

Query Match 71.4%; Score 30; DB 16; Length 321;
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9

:| |||||

Db 277 ERLVASKSF 285

RESULT 12

ID Q97JP2 PRELIMINARY; PRT; 331 AA.
 AC Q97JP2;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PREDICTED DEHYDROGENASE, YULF B.SUBTILIS ORTHOLOG.
 GN CAC1231.
 OS Clostridium acetobutylicum.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 OC Clostridium.
 OX NCBI_TaxID=1488;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
 RX MEDLINE=21359325; PubMed=11466286;
 RA Neelling J., Breton G., Onelchenko M.V., Makarova K.S., Zeng Q.,
 RA Gibson R., Lee H.M., Dubois J., Olu D., Hittl J., Soucaille P., Daly M.J.,
 RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Smith D.R.;
 RA Bennett G.N., Koonin E.V., Smith D.R.;
 RT "Genome sequence and comparative analysis of the solvent-producing
 RT bacterium Clostridium acetobutylicum.";
 RL J. Bacteriol. 183:4823-4838(2001).
 DR EMBL; AE007636; AAK79203.1; -;
 DR InterPro; IPR000683; GFO_IDH_MocA.
 DR Pfam; PF01408; GFO_IDH_MocA; 1.
 DR Complete proteome.
 KW Clostridium.
 SQ SEQUENCE 331 AA; 37659 MW; 8F89CB181A670C55 CRC64;

Query Match 71.4%; Score 30; DB 16; Length 331;
 Best Local Similarity 75.0%; Pred. No. 1.4e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8

:| |||||

Db 303 QRLKVSXS 310

RESULT 13

ID Q9M035 PRELIMINARY; PRT; 427 AA.
 AC Q9M035;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
 DE HYPOTHETICAL 48.1 KDA PROTEIN.
 GN T1008.80.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Pohl T., Weizenegger T., Bancroft I., Mewes H.W., Rudd S.,
 RA Lemcke K., Mayer K.F.X.;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RA Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL; AL161746; CAB81920.1; -;
 DR Hypothetical protein.

SQ SEQUENCE 427 AA; 48124 MW; 4B89FAADBBD33B0F CRC64;
Query Match 71.4%; Score 30; DB 10; Length 427;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 QRIKASKSF 9
:||:|||||
Db 140 RRVKASVSF 148
RESULT 14
Q926K2 ID Q926K2 PRELIMINARY; PRT; 478 AA.
AC Q926K2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PLI0058 PROTEIN.
GN PLI0058.
OS Listeria innocua.
OG Plasmid pLI100.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Listeria.
OX NCBI_TaxID=1642;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CLIP 11262 / SEROVAR 6A;
RX PubMed=11679669;
RA Glaser P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
RA Charbit A., Chetoui F., Couve E., de Daruvar A., Dehoux P.,
RA Damann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
RA Entian L.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
RA Jones L.-M., Kaerst U., Kretz J., Kuhn M., Kunst F., Kurapkat G.,
RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
RA Nordisiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
RA Remmel B., Rose M., Schluter T., Simoes N., Tierrez A.,
RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
RA "Comparative genomics of Listeria species.";
RL Science 294:849-852(2001).
DR EMBL; AL592102; CAC42056.1; -.
KW Plasmid; Complete proteome.
SQ SEQUENCE 478 AA; 56091 MW; 22E55927264B8E18 CRC64;
Query Match 71.4%; Score 30; DB 16; Length 478;
Best Local Similarity 66.7%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 QRIKASKSF 9
:||:|||||
Db 14 QRLKAKKF 22
RESULT 15
Q9FM26 ID Q9FM26 PRELIMINARY; PRT; 583 AA.
AC Q9FM26;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE SIMILARITY TO UNKNOWN PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=98290546; PubMed=9628582;
RA Sato S., Kaneko T., Kotani H., Nakamura Y., Asamizu E., Miyajima N.,

RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. IV.
RT Sequence features of the regions of 1,456,315 bp covered by nineteen
RT physically assigned pl and TAC clones.";
RL DNA Res. 5:41-54(1998).
DR EMBL; AB009052; BAB08544.1; -.
SQ SEQUENCE 583 AA; 66969 MW; BC423030D19F504B CRC64;
Query Match 71.4%; Score 30; DB 10; Length 583;
Best Local Similarity 75.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 RIKASKSF 9
:||:|||||
Db 268 RIKVSKTF 275
RESULT 16
Q932P5 ID Q932P5 PRELIMINARY; PRT; 583 AA.
AC Q932P5;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE AT5G40720/MNF13_240.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Cheuk R., Chen H., Kim C.J., Koesema E., Meyers M.C., Banh J.,
RA Bowsher J., Carninci P., Dale J.M., Goldsmith A.D., Hayashizaki Y.,
RA Ishida J., Jiang P.X., Jones T., Kamiya A., Karlin-Neumann G.,
RA Kawai J., Lam B., Lee J.M., Lin J., Liu S.X., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Pham P.K., Quach H.L., Sakurai T.,
RA Satou M., Seki M., Southwick A., Tang C.C., Toriumi M., Yamada K.,
RA Yamamura Y., Yu G., Yu S., Shinozaki K., Davis R.W., Theologis A.,
RA Ecker J.R.;
RA "Arabidopsis cDNA clones.";
RT Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY056405; AAL08261.1; -.
SQ SEQUENCE 583 AA; 66953 MW; 56839C2A10349F66 CRC64;
Query Match 71.4%; Score 30; DB 10; Length 583;
Best Local Similarity 75.0%; Pred. No. 2.5e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 2 RIKASKSF 9
:||:|||||
Db 268 RIKVSKTF 275
RESULT 17
Q15152 ID Q15152 PRELIMINARY; PRT; 726 AA.
AC Q15152; O00645;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PLAKOPHILIN
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Schmidt A., Hans H.W., Schaefer S., Nuber U.A., Zimbelmann R.,
RA Franke W.W.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.

RA Zimbelmann R.;
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Bosch A.;
 RL Submitted (SEP-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Schmidt A., Langbein L., Rode M., Praetzel S., Franke W.W.;
 RL J. Mol. Biol. 0:0-0(0).
 DR EMBL: 234974; CAA84426.1; -
 DR EMBL: 273678; CAA98022.1; -
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 5.
 DR SMART: SM00185; ARM; 3.
 DR PROSITE: PS0176; ARM_REPEAT; 3.
 SQ SEQUENCE 726 AA; 80496 MW; AE49B9396487AD31 CRC64;

Query Match 71.4%; Score 30; DB 4; Length 726;
 Best Local Similarity 75.0%; Pred. No. 3.2e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8
 I:||||:|
 Db 148 QRIKASRS 155

RESULT 18
 Q28161 PRELIMINARY; PRT; 727 AA.
 AC Q28161;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PLAKOPHILIN.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95237210; PubMed=7720719;
 RA Held H.W., Schmidt A., Zimbelmann R., Schaefer S., Franke W.W.,
 Nuber U.A.;
 RA "Desmosomes and cytoskeletal architecture in epithelial
 RT differentiation: cell type-specific plaque components and intermediate
 RT filament anchorage";
 RL Eur. J. Cell Biol. 65:229-245(1994).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95196971; PubMed=7890138;
 RA Held H.W., Schmidt A., Zimbelmann R., Schaefer S.,
 Winter-Simanowski S., Stump S., Keith M., Figge U., Schnolzer M.,
 Franke W.W.;
 RA "Cell type-specific desmosomal plaque proteins of the plakoglobin
 RT family: plakophilin 1 (band 6 protein).";
 RL Differentiation 58:113-131(1994).
 DR EMBL: 237975; CAA86029.1; -
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 5.
 DR SMART: SM00185; ARM; 4.
 DR PROSITE: PS0176; ARM_REPEAT; 3.
 SQ SEQUENCE 727 AA; 80180 MW; 3A27979279BCCEDF CRC64;

Query Match 71.4%; Score 30; DB 6; Length 727;
 Best Local Similarity 75.0%; Pred. No. 3.2e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8
 I:||||:|
 Db 149 QRIKASRS 156

RESULT 19
 P97350 PRELIMINARY; PRT; 728 AA.
 ID P97350;
 AC P97350;
 DT 01-MAY-1997 (TrEMBLrel. 03, Created)
 DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PLAKOPHILIN 1.
 GN PKP1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57 BLAK/6; TISSUE=SKIN;
 RA Nimrich V., Hunziker A.H., Franke W.W.;
 RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 DR EMBL: Y07941; CAA69240.1; -
 DR MGD: MGI:1328359; Pkp1.
 DR InterPro: IPR000225; Armadillo.
 DR Pfam: PF00514; Armadillo_seg; 5.
 DR SMART: SM00185; ARM; 3.
 DR PROSITE: PS0176; ARM_REPEAT; 3.
 SQ SEQUENCE 728 AA; 80896 MW; BDAC5BA7B4118AC0 CRC64;

Query Match 71.4%; Score 30; DB 11; Length 728;
 Best Local Similarity 75.0%; Pred. No. 3.2e+02;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8
 I:||||:|
 Db 149 QRIKASRS 156

RESULT 20
 Q96825 PRELIMINARY; PRT; 826 AA.
 ID Q96825;
 AC Q96825;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
 DE PUTATIVE GDP-GTP EXCHANGE FACTOR.
 GN EXC-5
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Suzuki N., Buechner M., Nishiwaki K., Hall D.H., Nakanishi H.,
 Takai Y., Hisamoto N., Matsumoto K.;
 RA "A putative GDP-GTP exchange factor is required for development of the
 RT excretory cell in C. elegans.";
 RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AB060647; BAB43906.1; -
 SQ SEQUENCE 826 AA; 94302 MW; 3810677CEB7684A9 CRC64;

Query Match 71.4%; Score 30; DB 5; Length 826;
 Best Local Similarity 62.5%; Pred. No. 3.6e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 2 RIKASKSF 9
 I:||||:
 Db 202 RVKSSKSY 209

RESULT 21
 Q18372 PRELIMINARY; PRT; 875 AA.
 ID Q18372;
 AC Q18372;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE C33D9.1 PROTEIN.
 GN C33D9.1
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Lloyd C.R.;
 RL Submitted (DEC-1995) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99069613; PubMed=9851916;
 RA none;
 RT "Genome sequence of the nematode C.elegans: A platform for
 RT investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL; Z68159; CAA92282.1; -.
 DR InterPro: IPR001849; PH.
 DR InterPro: IPR000219; RhoGEF.
 DR InterPro: IPR000306; Znf_FYVE.
 DR Pfam; PF00169; PH; 2.
 DR Pfam; PF00621; RhoGEF; 1.
 DR SMART; SM00064; FYVE; 1.
 DR SMART; SM00233; PH; 2.
 DR SMART; SM00325; RhoGEF; 1.
 DR PROSITE; PS50003; PH_DOMAIN; 1.
 SQ SEQUENCE 875 AA; 100290 MW; 2C7DEALE2C25852 CRC64;
 Query Match 71.4%; Score 30; DB 5; Length 875;
 Best Local Similarity 62.5%; Pred. No. 3.8e+02;
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 2 RIKASKSF 9
 ID 233 RVRKRSY 240
 ID Q9PFT8 PRELIMINARY; PRT; 985 AA.
 AC Q9PFT8;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PYRUVATE DEHYDROGENASE.
 GN XF0669.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G.; Reinach F.C.; Arruda P.; Abreu F.A.; Acencio M.;
 RA Alvarenga R.; Alves L.M.C.; Arya J.E.; Baia G.S.; Baptista C.S.;
 RA Barros M.H.; Bonaccorsi E.D.; Bordin S.; Bove J.M.; Briones M.R.S.;
 RA Bueno M.R.P.; Camargo A.A.; Camargo L.E.A.; Carraro D.M.; Carrier H.;
 RA Colauto N.B.; Colombo C.; Costa F.F.; Costa M.C.R.; Costa-Neto C.M.;
 RA Coutinho L.L.; Cristofani M.; Dias-Neto E.; Docena C.; El-Dorry H.;
 RA Pacincai A.P.; Ferreira A.J.S.; Ferreira V.C.A.; Ferro J.A.;
 RA Fraga J.S.; Franca S.C.; Franco M.C.; Frohme M.; Furian L.R.;
 RA Garnier M.; Goldman G.H.; Goldman M.H.S.; Gomes S.L.; Gruber A.;
 RA Ho P.L.; Hoheisel J.D.; Junqueira M.L.; Kemper E.L.; Kitajima J.P.;
 RA Krieger J.E.; Kuramae E.E.; Laigret F.; Lambais M.R.; Leite L.C.C.;
 RA Lemos E.G.M.; Lemos M.V.F.; Lopes S.A.; Lopes C.R.; Machado J.A.;
 RA Machado M.A.; Madeira A.N.B.N.; Madeira H.M.F.; Marino C.L.;
 RA Marques M.V.; Martins E.A.L.; Martins E.M.F.; Matsukuma A.Y.;
 RA Menck C.F.M.; Miracca E.C.; Miyaki C.Y.; Monteiro-Vitorello C.B.;
 RA Moon D.H.; Nagai M.A.; Nascimento A.L.T.O.; Netto L.E.S.

RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
 RA Peixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pesquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
 RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
 RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsubako M.H.,
 RA Vallada H., Van Sluys M.A., Verjovsky-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa."
 RL Nature 406:151-159(2000).
 DR EMBL; AE003911; AAF83479.1; -.
 KW Complete proteome.
 SQ SEQUENCE 985 AA; 110224 MW; 658C284A9A3C3CD6 CRC64;
 Query Match 71.4%; Score 30; DB 16; Length 985;
 Best Local Similarity 66.7%; Pred. No. 4.3e+02;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ORIKASKSF 9
 ID 560 RRTKSKSF 568
 ID Q9ZVX1 PRELIMINARY; PRT; 1102 AA.
 AC Q9ZVX1;
 DT 01-MAY-1999 (TREMBlrel. 10, Created)
 DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE PUTATIVE UBIQUITIN-CONJUGATING ENZYME.
 GN AT2G16920.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RX MEDLINE=20083487; PubMed=10617197;
 RA Lin X.; Kaul S.; Rounsley S.D.; Shea T.P.; Benito M.-I.; Town C.D.;
 RA Fujii C.Y.; Mason T.M.; Bowman C.L.; Barnstead M.E.; Feldblum T.V.;
 RA Buell C.R.; Ketchum K.A.; Lee J.J.; Ronning C.M.; Roo H.; Moffat K.S.;
 RA Cronin L.A.; Shen M.; VanAken S.E.; Umayam L.; Tallon L.J.; Gill J.E.;
 RA Adams M.D.; Carrera A.J.; Creasy T.H.; Goodman H.M.; Somerville C.R.;
 RA Copenhaver G.P.; Preuss D.; Nierman W.C.; White O.; Eisen J.A.;
 RA Salzberg S.L.; Fraser C.M.; Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
 RT thaliana."
 RL Nature 402:761-768(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: CATALYZES THE COVALENT ATTACHMENT OF UBIQUITIN TO OTHER
 CC PROTEINS (BY SIMILARITY).
 CC -!- CATALYTIC ACTIVITY: ATP + UBIQUITIN + PROTEIN LYSINE = AMP +
 CC PYROPHOSPHATE + PROTEIN N-UBIQUITYLLYSINE.
 CC -!- PATHWAY: SECOND STEP IN UBIQUITIN CONJUGATION.
 CC -!- MISCELLANEOUS: A CYSTEINE RESIDUE IS REQUIRED FOR UBIQUITIN-
 CC THIOLESTER FORMATION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE UBIQUITIN-CONJUGATING ENZYME FAMILY.
 DR EMBL; AC005167; G64223.1; -.
 DR HSP; 002159; 2UCZ.
 DR InterPro: IPR000169; Thiolprot_act_site.
 DR InterPro: IPR000608; UBQ_conjugat.
 DR Pfam; PF00179; UQ_con; 1.
 DR SMART; SM00212; UBCC; 1.
 DR PROSITE; PS00639; THIOL_PROTEASE_HIS; UNKNOWN_1.

DR PROSITE; PS00127; UBIQUITIN_CONJUGAT_2; 1.
 KW Ligase: Ubiquitin conjugation.
 SQ SEQUENCE 1102 AA; 122182 MW; DA2C11385355E8D9 CRC64;

Query Match 71.4%; Score 30; DB 10; Length 1102;
 Best Local Similarity 75.08; Pred. No. 4.8e+02;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
 II:III:II
 DB 257 QRVKASS 264

RESULT 24

P89914 ID P89914 PRELIMINARY; PRT; 140 AA.
 AC P89914;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE 16 KDA PROTEIN (FRAGMENT).
 OS Tobacco rattle virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobravirus.
 OX NCBI_TaxID=12295;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SYRAIN-CANADIAN (CAN);
 RA Kawchuk L.M., Lynch D.R., Leggett F.L., Howard R.J., McDonald J.G.;
 RT "Characterization of a Canadian Tobacco Rattle Virus Isolate."
 RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
 FR EMBL; U63727; AAB48382.1; -.
 FT NON_TER 1
 SQ SEQUENCE 140 AA; 16274 MW; 33B124A3F8A90040 CRC64;

Query Match 59.0%; Score 29; DB 12; Length 140;
 Best Local Similarity 55.8%; Pred. No. 1e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
 II:III:II
 DB 89 ERIRAEKTF 97

RESULT 25

O00897 ID O00897 PRELIMINARY; PRT; 149 AA.
 AC O00897;
 DT 01-JUL-1997 (TREMBLrel. 04, Created)
 DT 01-JUL-1997 (TREMBLrel. 04, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE CALMODULIN-LIKE PROTEIN.
 GN CALB.
 OS Dictyostelium discoideum (Slime mold).
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
 OX NCBI_TaxID=44689;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-AX3;
 RX MEDLINE-20279912; PubMed=10818233;
 RA Rosel D., Puta F., Blahusova A., Smykal P., Folk P.;
 RT "Molecular characterization of a calmodulin-like Dictyostelium protein CalB."
 RL FEBS Lett. 473:323-327(2000).
 CC -1- SIMILARITY: TO OTHER EF-HAND CALCIUM BINDING PROTEINS.
 DR EMBL; AF001981; AAB60882.1; -.
 DR HSP; P02593; IAK8.
 DR InterPro; IPR002048; EF-hand.
 DR Pfam; PF00036; ehand; 4.
 DR SMART; SM00054; EFh; 3.
 DR PROSITE; PS00018; EF_HAND; UNKNOWN_1.
 KW Calcium-binding.
 SQ SEQUENCE 149 AA; 16827 MW; D7ADD31A83EA422D CRC64;

Query Match 69.0%; Score 29; DB 5; Length 149;
 Best Local Similarity 85.7%; Pred. No. 1.1e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IKAKSKSF 9
 II:III:II
 DB 142 IKSKSF 148

RESULT 26

Q9PMA5 ID Q9PMA5 PRELIMINARY; PRT; 172 AA.
 AC Q9PMA5;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE NADH DEHYDROGENASE I CHAIN J (EC 1.6.5.3).
 GN NUOJ OR CJ1570.
 OS Campylobacter jejuni.
 OC Bacteria; Proteobacteria; epsilon subdivision; Campylobacter group;
 OC Campylobacter.
 OX NCBI_TaxID=197;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-NCTC 11168;
 RX MEDLINE-20150912; PubMed=10688204;
 RA Parkhill J., Wren B.W., Mungall K., Kettle J.M., Churcher C.,
 RA Basham D., Chillingworth T., Davies R.M., Feltwell T., Holtroyd S.,
 RA Jais K., Karlyshev A.V., Moule S., Pallen M.J., Penn C.W.,
 RA Quail M.A., Rajandream M.A., Rutherford K.M., van Vliet A.H.M.,
 RA Whitehead S., Barrell B.G.;
 RT "The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences."
 RL Nature 403:665-668(2000).
 DR EMBL; ALI39079; CAB73538.1; -.
 DR InterPro; IPR001457; Oxidored_q3.
 DR Pfam; PF00499; oxidored_q3; 1.
 KW Complete proteome.
 SQ SEQUENCE 172 AA; 19191 MW; D0FCF1A4F5B8F0E2 CRC64;

Query Match 69.0%; Score 29; DB 16; Length 172;
 Best Local Similarity 55.6%; Pred. No. 1.2e+02;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
 III:II:II
 DB 84 EOLKAKSKF 92

RESULT 27

Q94JH8 ID Q94JH8 PRELIMINARY; PRT; 175 AA.
 AC Q94JH8;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE P0638D12.11 PROTEIN.
 GN P0638D12.11.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoideae; Oryzaceae; Oryza.
 OX NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CV. NIPPONBARE;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC clone:p0638D12."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AP002972; BAB55503.1; -.
 SQ SEQUENCE 175 AA; 19328 MW; 514C88224F275BBF CRC64;

Query Match 69.0%; Score 29; DB 10; Length 175;
Best Local Similarity 55.6%; Pred. No. 1.3e+02;
Matches 5; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

Db 111 KRIKSKSF 119

RESULT 28

Q9VGZ7 PRELIMINARY; PRT; 182 AA.

AC Q9VGZ7;

DT 01-MAY-2000 (TReMBLrel. 13, Created)

DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)

DE CG18545 PROTEIN.

GN CG18545.

OS Drosophila melanogaster (fruit fly).

OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;

OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;

OC Ephydroidea; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;

RX MEDLINE=20196006; PubMed=10731132;

RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,

RA Ananidis P.G., Scher S.E., Li P.W., Hoskins R.A., Gallie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,

RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,

RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,

RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,

RA Arkil J.F., Agayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,

RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,

RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,

RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,

RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,

RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

RA Foslter C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,

RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,

RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,

RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,

RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,

RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,

RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,

RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,

RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,

RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith H.,

RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,

RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,

RA Wang Z.-Y., Wasman D.A., Weinstein G.M., Weissenbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,

RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,

RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

RT "The genome sequence of Drosophila melanogaster.";

RL Science 287:2185-2195(2000).

DR EMBL; AE003686; AAF5424.1; -

DR FlyBase; FBgn0037812; CG18545.

SQ SEQUENCE 182 AA; 20774 MW; A0AAB88FE90ABD22 CRC64;

Query Match 69.0%; Score 29; DB 5; Length 182;

Best Local Similarity 66.7%; Pred. No. 1.3e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

Db 61 KRIKLEKSF 69

RESULT 29

Q91BI4

ID Q91BI4 PRELIMINARY; PRT; 201 AA.

AC Q91BI4;

DT 01-DEC-2001 (TReMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE LATE EXPRESSION FACTOR 12.

OS Spodoptera litura nucleopolyhedrovirus.

OC Viruses; dsDNA viruses, no RNA stage; Baculoviridae;

OC Nucleopolyhedrovirus.

OX NCBI_TaxID=46242;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=G2;

RX MEDLINE=21425398; PubMed=11511416;

RA Pang Y., Yu J., Wang L., Hu X., Bao W., Li G., Chen C., Han H., Hu S.,

RA Yang H.;

RT "Sequence Analysis of the Spodoptera litura Multicapsid

RT Nucleopolyhedrovirus Genome.";

RL Virology 287:391-404(2001).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=G2;

RA Yu J., Wang L., Hu X., Pang Y.;

RL Submitted (DEC-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF325155; AAL01724.1;

SQ SEQUENCE 201 AA; 23598 MW; 79E3A3FD904F3C4C CRC64;

Query Match 69.0%; Score 29; DB 12; Length 201;

Best Local Similarity 75.0%; Pred. No. 1.5e+02;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKASKSF 9

Db 72 RIKCESF 79

RESULT 30

ID Q95KC4

ID Q95KC4 PRELIMINARY; PRT; 202 AA.

AC Q95KC4;

DT 01-DEC-2001 (TReMBLrel. 19, Created)

DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)

DE HYPOTHETICAL 20.7 KDA PROTEIN.

OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;

OC Cercopitheciinae; Macaca.

OX NCBI_TaxID=9541;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=MEDULLA OBLOGATA;

RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirai M., Terao K.,

RA Suzuki Y., Sugano S., Hashimoto K.;

RT "Isolation of full-length cDNA clones from macaque brain CDNA

RT libraries.";

RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB062980; BAB60746.1; -

KW Hypothetical protein.

SQ SEQUENCE 202 AA; 20733 MW; 883C313D6C58E5A4 CRC64;

Query Match 69.0%; Score 29; DB 6; Length 202;

Best Local Similarity 66.7%; Pred. No. 1.5e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

||||:|:|
Db 188 QRIKEAKRF 196

RESULT 31

Q94A31 PRELIMINARY; PRT; 236 AA.
AC Q94A31;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE AT5G02520/T22P11.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,
RA Banh J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,
RA Shinzaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis cDNA clones";
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY050424; AAK91440.1; -
SQ SEQUENCE 236 AA; 26354 MW; 93EB9F220EE9FF8 CRC64;

Query Match 69.0%; Score 29; DB 10; Length 236;

Best Local Similarity 62.5%; Pred. No. 1.7e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 RIKASKSF 9

||||:|:|
Db 94 RVKVSFAF 101

RESULT 32

O42159 PRELIMINARY; PRT; 244 AA.
AC O42159;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TRYPSINOGEN B1 PRECURSOR (FRAGMENT).
GN TRYPB1.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RA Roach J.C.;
RT "The Molecular Evolution of the Vertebrate Trypsinogens.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL: AF011900; AAB69656.1; -
DR HSSP; P00763; IDPO.
DR MEROPS; S01.151; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT SIGNAL <1 13
FT CHAIN 14 245
SQ SEQUENCE 245 AA; 26001 MW; 9A932508B896C93E CRC64;

Query Match 69.0%; Score 29; DB 13; Length 245;

Best Local Similarity 75.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8

||||:|:|
Db 86 QRIQASKA 93

RESULT 34

Q9HNY9 PRELIMINARY; PRT; 268 AA.
AC Q9HNY9;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE QUINOLINATE PHOSPHORIBOSYLTRANSFERASE.
GN NADC OR VNG1884G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;

FT NON_TER 1
FT SIGNAL <1 12 POTENTIAL.
FT CHAIN 13 244 TRYPSIN B1.
SQ SEQUENCE 244 AA; 25903 MW; C4582EE07E3B8007 CRC64;

Query Match 69.0%; Score 29; DB 13; Length 244;

Best Local Similarity 75.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8

||||:|:|
Db 85 QRIQASKA 92

RESULT 33

O42160 PRELIMINARY; PRT; 245 AA.
AC O42160;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE TRYPSINOGEN B2 PRECURSOR (FRAGMENT).
GN TRYPB2.
OS Petromyzon marinus (Sea lamprey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Hyperoartia;
OC Petromyzontiformes; Petromyzontidae; Petromyzon.
OX NCBI_TaxID=7757;
RN [1]
RP SEQUENCE FROM N.A.
RA Roach J.C.;
RT "The Molecular Evolution of the Vertebrate Trypsinogens.";
RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
TRYPSIN FAMILY.
DR EMBL: AF011901; AAB69657.1; -
DR HSSP; P00763; IDPO.
DR MEROPS; S01.151; -
DR InterPro; IPR001314; Chymotrypsin.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00722; CHYMOTRYPSIN.
DR SMART; SM00020; Tryp_SPC; 1.
DR PROSITE; PS02040; TRYPSIN_DOM; 1.
DR PROSITE; PS00134; TRYPSIN_HIS; UNKNOWN_1.
DR PROSITE; PS00135; TRYPSIN_SER; 1.
KW Hydrolase; Serine protease; Signal.
FT NON_TER 1
FT SIGNAL <1 13
FT CHAIN 14 245
SQ SEQUENCE 245 AA; 26001 MW; 9A932508B896C93E CRC64;

Query Match 69.0%; Score 29; DB 13; Length 245;

Best Local Similarity 75.0%; Pred. No. 1.8e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8

||||:|:|
Db 86 QRIQASKA 93

```
RN SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dasarma S.;
RT "Genome sequence of Halobacterium species NRC-1."
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE005087; AAG20081.1; -.
DR HSSP; P30012; IQAP.
DR InterPro; IPR002638; ORPase.
DR Pfam; PF01729; ORPase; 1.
DR Pfam; PF02749; ORPase_N; 1.
DR Transferase; Glycosyltransferase; Complete proteome.
SQ SEQUENCE 268 AA; 27635 MW; BB8A6E6C64C67EFF CRC64;

Query Match 69.0%; Score 29; DB 17; Length 268;
Best Local Similarity 55.6%; Pred. No. 1.9e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
   :|:|:|
Db 172 ERFRAKSF 180

RESULT 35
Q9VJG8 PRELIMINARY; PRT; 308 AA.
AC Q9VJG8;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE CG6012 PROTEIN.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celnik S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Moberly B., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
```

```
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RL Science 287:2185-2195(2000).
CC -1- SIMILARITY: BELONGS TO THE SHORT-CHAIN DEHYDROGENASES/REDUCTASES
   (SDR) FAMILY.
CC EMBL; AE003652; AAF53581.1; -.
DR FlyBase; FBgn0032615; CG6012.
DR InterPro; IPR002198; ADH_short.
DR Pfam; PF00106; adh_short; 1.
DR PRINTS; PR00080; SDRFAMILY.
KW Oxidoreductase.
SQ SEQUENCE 308 AA; 34163 MW; 1A4F208AF6FCE04 CRC64;

Query Match 69.0%; Score 29; DB 5; Length 308;
Best Local Similarity 85.7%; Pred. No. 2.2e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASK 7
   |::|||
Db 171 QRMKASK 177

RESULT 36
Q9GRJ8 PRELIMINARY; PRT; 325 AA.
ID Q9GRJ8;
AC Q9GRJ8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL 35.1 KDA PROTEIN.
DE P214.44.
OS Leishmania major.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.
OX NCBI_TaxID=5664;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RA Saunders D., Murphy L., Harris D., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FRIEDLIN;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagherzadeh A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome."
RL Genome Res. 8:135-145(1998).
DR EMBL; AL449144; CAC14651.1; -.
KW Hypothetical protein.
SQ SEQUENCE 325 AA; 35104 MW; 62D04B2B77001BC8 CRC64;

Query Match 69.0%; Score 29; DB 5; Length 325;
Best Local Similarity 55.6%; Pred. No. 2.4e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
   :|:|:|
Db 138 ERLKAEQSF 146

RESULT 37
Q07224
```

```
ID Q07224 PRELIMINARY; PRT; 368 AA.
AC Q07224;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HYPOTHETICAL PERMEASE (FRAGMENT).
GN YOR267C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=97051594; PubMed=8896271;
RA Charet G., Bernardi A., Sor F.J.;
RT "DNA sequence analysis of the VP1-SNF2 region on chromosome XV of
RT Saccharomyces cerevisiae.";
RL Yeast 12:1059-1064(1996).
RN [2]
RP SEQUENCE FROM N.A.
RA Jauniaux J.C., Poirey R.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Charet G., Sor F.;
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; X89633; CAA61794.1; -
DR EMBL; Z75176; CAA99492.1; -
DR SGD; S0005793; YOR267C.
DR InterPro; IPR0010194; Atpase_alpha_beta.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PROSITE; PS00152; ATPASE_ALPHA_BETA; UNKNOWN_1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Serine/threonine-protein kinase; Transferase.
FT NON_TER 368
SQ SEQUENCE 368 AA; 41341 MW; 512FB37C2BB11574 CRC64;

Query Match 69.0%; Score 29; DB 3; Length 368;
Best Local Similarity 66.7%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
Db 30 QLIRSSKSF 38
1:|:|:|:|:|

RESULT 38
O62258 PRELIMINARY; PRT; 368 AA.
AC O62258;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE F53F8.5 PROTEIN.
GN F53F8.5.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA Matthews L.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
```

```
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL; Z81547; CAB04458.1; -
DR InterPro; IPR001660; SAM.
DR Pfam; PF00536; SAM; 1.
DR SMART; SM00454; SAM; 1.
SQ SEQUENCE 368 AA; 40392 MW; 6BF20077F7C6749B CRC64;

Query Match 69.0%; Score 29; DB 5; Length 368;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8
Db 59 RRIKAAS 66
1:|:|:|:|:|

RESULT 39
O9A119 PRELIMINARY; PRT; 439 AA.
AC O9A119;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE PUTATIVE SUGAR TRANSPORTER SUGAR BINDING LIPOPROTEIN.
GN SPY0252.
OS Streptococcus pyogenes.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1314;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SF370 / ATCC 700294 / SEROTYPE M1;
RX MEDLINE=21192684; PubMed=11296296;
RA Ferretti J.J., McShan W.M., Ajdic D.J., Savic D.J., Savic G., Lyon K.,
RA Primeaux C., Sezate S., Suvorov A.N., Kenton S., Lai H.S., Lin S.P.,
RA Qian Y., Jia H.G., Najjar F.Z., Ren Q., Zhu H., Song L., White J.,
RA Yuan X., Clifton S.W., Roe B.A., McLaughlin R.;
RT "Complete genome sequence of an M1 strain of Streptococcus pyogenes.";
RL PROC. Natl. Acad. Sci. U.S.A. 98:4658-4663(2001).
DR EMBL; AE006492; AAK3328.1; -
KW Sugar transport; Lipoprotein; Complete proteome.
SQ SEQUENCE 439 AA; 48506 MW; 4541C029508748AC CRC64;

Query Match 69.0%; Score 29; DB 16; Length 439;
Best Local Similarity 55.6%; Pred. No. 3.2e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
Db 335 QKVAASKTF 343
1:|:|:|:|:|

RESULT 40
Q97E56 PRELIMINARY; PRT; 463 AA.
AC Q97E56;
DT 01-OCT-2001 (TrEMBLrel. 18, Created)
DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ASPARTYL/ASPARAGINYL-TRNA SYNTHETASE.
GN CAC3260.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
OC Clostridium.
OX NCBI_TaxID=1488;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 824 / DSM 792 / VKM B-1787;
RX MEDLINE=21359325; PubMed=11466286;
RA Noelling J., Breton G., Omeichenko M.V., Makarova K.S., Zeng Q.,
RA Gibson R., Lee H.M., Dubois J., Qiu D., Hitti J., Wolf Y.I.,
```

RA Tatusov R.L., Sabathe F., Doucette-Stamm L., Soucaille P., Daly M.J.,
RA Bennett G.N., Koonin E.V., Smith D.R.;
RT "Genome sequence and comparative analysis of the solvent-producing
RT bacterium Clostridium acetobutylicum";
RL J. Bacteriol. 183:4823-4838(2001).
DR EMBL; AE007821; AAK81194.1; -
DR InterPro; IPR002106; AA_TRNA_ligase_II.
DR InterPro; IPR004364; tRNA-synt_2.
DR InterPro; IPR002312; tRNA-synt_2.
DR InterPro; IPR004365; tRNA-anti.
DR Pfam; PF00152; tRNA-synt_2; 1.
DR Pfam; PF01336; tRNA-anti; 1.
DR PRINTS; PR01042; TRNASYNTHASP.
DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; UNKNOWN_1.
DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; UNKNOWN_1.
KW Aminoacyl-tRNA synthetase; Complete proteome.
SQ SEQUENCE 463 AA; 53439 MW; BC70BD19704ADB43 CRC64;

Query Match 69.0%; Score 29; DB 16; Length 463;
Best Local Similarity 85.7%; Pred. No. 3.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IRASKSF 9
DB 30 IRASKSF 36

RESULT 41
Q9L0P9 PRELIMINARY; PRT; 468 AA.
AC Q9L0P9;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE PUTATIVE MEMBRANE PROTEIN.
GN SCD8A.15C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Seeger K.J., Harris D.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdeno A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapaite D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL160331; CAB77336.1; -
SQ SEQUENCE 468 AA; 48544 MW; C0FD487AA341CAC0 CRC64;

Query Match 69.0%; Score 29; DB 2; Length 468;
Best Local Similarity 85.7%; Pred. No. 3.4e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IRASKSF 9
DB 327 IRASKSF 333

RESULT 42
O58616 PRELIMINARY; PRT; 526 AA.
AC O58616;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 60.1 KDA PROTEIN PH0917.
GN Pyrococcus horikoshii.
OS Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudo H., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -!- SIMILARITY: TO DEAD/DEAH BOX HELICASE FAMILY.
DR EMBL; AP000004; BAA30013.1; -
DR InterPro; IPR001410; DEAD.
DR InterPro; IPR001650; Helicase_C.
DR Pfam; PF00270; DEAD; 1.
DR Pfam; PF00271; helicase_C; 1.
DR SMART; SM00490; HELTCC; 1.
KW ATP-binding; Complete proteome; Helicase; Hypothetical protein.
SQ SEQUENCE 526 AA; 60145 MW; 6B5DB65BFCADF6F5 CRC64;

Query Match 69.0%; Score 29; DB 17; Length 526;
Best Local Similarity 75.0%; Pred. No. 3.8e+02;
Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RIKASKSF 9
DB 144 RIKASKSF 151

RESULT 43
Q9PPZ1 PRELIMINARY; PRT; 563 AA.
AC Q9PPZ1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)
DE HYPOTHETICAL PROTEIN UU499.
GN UU499.
OS Ureaplasma parvum (Ureaplasma urealyticum biotype 1).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Ureaplasma.
OX NCBI_TaxID=134821;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SEROVAR 3;
RX MEDLINE=20500219; PubMed=11048724;
RA Glass J.I., Lefkowitz E.J., Glass J.S., Heiner C.R., Chen E.Y.,
RA Cassell G.H.;
RT "The complete sequence of the mucosal pathogen Ureaplasma
RT urealyticum";
RL Nature 407:757-762(2000).
DR EMBL; AF002148; AAF30911.1; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 563 AA; 67786 MW; 7A8103BA3B3C052D CRC64;

Query Match 69.0%; Score 29; DB 16; Length 563;
Best Local Similarity 66.7%; Pred. No. 4.1e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

```

Db 459 QKINASNSF 467
RESULT 44
ID Q9LZ50 PRELIMINARY; PRT; 584 AA.
AC Q9LZ50;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL 64.4 KDA PROTEIN.
GN T22P11.110.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,
RA Bancroft I., Mewes H.W., Rudd S., Lemcke K., Mayer K.F.X.;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA EU Arabidopsis sequencing project;
RL Submitted (APR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL162971; CAB85989.1; -.
KW Hypothetical protein.
SQ SEQUENCE 584 AA; 64386 MW; 7E23365AD92A22CD CRC64;

Query Match 69.08; Score 29; DB 10; Length 584;
Best Local Similarity 62.58; Pred. No. 4.3e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 2 RIKASKSF 9
I:| |||
Db 456 RVKVSFAC 463

RESULT 45
ID Q9VG28 PRELIMINARY; PRT; 669 AA.
AC Q9VG28;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CG12592 PROTEIN.
GN CG12592.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananaitides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiril J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S.,
RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
RA Burtis K.C., Busam D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,

```

```

RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003686; AAF54523.2; -.
DR FlyBase; FBgn0037810; CG12819.
SQ SEQUENCE 669 AA; 74855 MW; 167F7A784C3CFA52 CRC64;

Query Match 69.08; Score 29; DB 5; Length 669;
Best Local Similarity 66.78; Pred. No. 4.9e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
I:| |||
Db 61 KRIRKLSKF 69

Search completed: November 6, 2002, 12:12:18
Job time : 21.4444 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 5.33333 Seconds
(without alignments)
65.339 Million cell updates/sec

Title: US-09-905-083-99
Perfect score: 42
Sequence: 1 ORIKASKSF 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues
Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	42	100.0	253	1 KLK7_HUMAN	P49862 homo sapien
2	33	78.6	314	1 RLAO_THEAC	P57692 thermoplasm
3	33	78.6	427	1 KAIN_HUMAN	P29622 homo sapien
4	31	73.8	247	1 PCPC_FLAS3	Q03520 flavobacter
5	31	73.8	298	1 SNAB_BOVIN	P81126 bos taurus
6	31	73.8	298	1 SNAB_HUMAN	Q9h115 homo sapien
7	31	73.8	3038	1 TRIO_HUMAN	O75962 homo sapien
8	30	71.4	142	1 BET1_YEAST	P22804 saccharomyc
9	30	71.4	517	1 YB3C_SCHPO	O14342 schizosacch
10	30	71.4	626	1 PEPO_LACLA	Q09145 lactococcus
11	30	71.4	626	1 PEPO_LACLC	Q09145 lactococcus
12	30	71.4	727	1 PKF1_BOVIN	Q28161 bos taurus
13	30	71.4	728	1 PKP1_MOUSE	P97350 mus musculus
14	30	71.4	747	1 PKP1_HUMAN	Q13835 homo sapien
15	30	71.4	3660	1 DMD_CHICK	P11533 gallus gall
16	29	69.0	141	1 V16K_TRVPL	P33777 tobacco rat
17	29	69.0	182	1 ORN_HAERIN	P45340 haemophilus
18	29	69.0	463	1 SYN_CLOAB	Q97656 clostridium
19	29	69.0	484	1 6PGD_ACTAC	P70718 actinobacil
20	29	69.0	588	1 CALI_HUMAN	Q13939 homo sapien
21	29	69.0	853	1 DIA3_HUMAN	Q9nsv4 homo sapien
22	29	69.0	1018	1 YK22_CAEEL	P41993 caenorhabdi
23	28	66.7	46	1 DIUH_LOEMI	P23465 locusta mig
24	28	66.7	61	1 CSRA_ECOLI	P31803 escherichia
25	28	66.7	61	1 CSRA_ERWCA	Q47620 erwinia car
26	28	66.7	69	1 CSRA_SERMA	O85735 serratia ma
27	28	66.7	156	1 GREB_BACSD	P80240 bacillus su
28	28	66.7	158	1 GREB_BACHD	Q9kd47 bacillus ha
29	28	66.7	306	1 MK16_YEAST	P10962 saccharomyc
30	28	66.7	356	1 MURB_BUCAI	P57153 buchmera ap
31	28	66.7	415	1 PGK_YEAST	P00560 saccharomyc
32	28	66.7	418	1 YCCE_ECOLI	P36661 escherichia
33	28	66.7	439	1 YM3M_CAEEL	P91343 caenorhabdi

RESULT 1

ID	KLK7_HUMAN	STANDARD;	PRT;	253 AA.
AC	P49862:			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	01-OCT-1996 (Rel. 34, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Kallikrein 7 precursor (EC 3.4.21.-) (Stratum corneum chymotryptic enzyme) (hSCCE).			
DE	enzyme) (hSCCE).			
GN	KLK7 OR PRSS6 OR SCCE.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 23-53.			
RC	TISSUE=Skin;			
RX	MEDLINE=94308225; PubMed=8034709;			
RA	Hansson L., Stroemqvist M., Baeckman A., Wallbrandt P., Carlstein A., Egelrud T.			
RT	"Cloning, expression, and characterization of stratum corneum chymotryptic enzyme. A skin-specific human serine proteinase.";			
RL	J. Biol. Chem. 269:19420-19426(1994).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Keratinocytes;			
RA	Yousef G.M., Scorillas A., Diamandis E.P.;			
RT	"Molecular characterization, mapping and tissue expression of the human stratum corneum chymotryptic enzyme gene.";			
RL	Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20510030; PubMed=11054574;			
RA	Gan L., Lee I., Smith R., Argonza-Barrett R., Lei H., McCuaig J., Moss P., Paepfer B., Wang K.;			
RT	"Sequencing and expression analysis of the serine protease gene cluster located in chromosome 19q13 region.";			
RL	Gene 257:119-130(2000).			
RN	[4]			
RP	CHARACTERIZATION.			
RX	MEDLINE=95314630; PubMed=7794273;			
RA	Skytt A., Stroemqvist M., Egelrud T.;			
RT	"Primary substrate specificity of recombinant human stratum corneum chymotryptic enzyme.";			
RL	Biochem. Biophys. Res. Commun. 211:586-589(1995).			
CC	-!- FUNCTION: MAY CATALYZE THE DEGRADATION OF INTERCELLULAR COHESIVE STRUCTURES IN THE CORNIFIED LAYER OF THE SKIN IN THE CONTINUOUS SHEDDING OF CELLS FROM THE SKIN SURFACE. SPECIFIC FOR AMINO ACID RESIDUES WITH AROMATIC SIDE CHAINS IN THE P1 POSITION. SCCE CLEAVES INSULIN B CHAIN AT 6-LEU- -CYS-7, 16-TYR- -LEU-17, 25-PHE- -TYR-26, AND 26-TYR- -THR-27. COULD PLAY A ROLE IN THE ACTIVATION OF PRECURSORS TO INFLAMMATORY CYTOKINES			
CC	-!- TISSUE SPECIFICITY: IT IS ABUNDANTLY EXPRESSED IN THE SKIN AND IS EXPRESSED BY KERATINOCYTES IN THE EPIDERMIS. VERY LOW LEVELS ARE ALSO SEEN IN THE BRAIN AND KIDNEY.			
CC	-!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE			

Q9chu6 lactococcus
P96789 lactococcus
P53061 saccharomyc
P78032 mycoplasma
O15265 homo sapien
Q48509 lactobacill
Q929k5 bacillus ha
P12874 bacillus su
Q93978 mus musculus
P09c173 lactococcus
P06324 mus musculus
P01740 mus musculus

ALIGNMENTS

34 28 66.7 472 1 6PGD_LACLA
35 28 66.7 472 1 6PGD_LACLC
36 28 66.7 704 1 YG29_YEAST
37 28 66.7 711 1 TOP1_MYCPN
38 28 66.7 892 1 ATX7_HUMAN
39 27 64.3 62 1 LAFX_LACJO
40 27 64.3 86 1 RS17_BACHD
41 27 64.3 86 1 RS17_BACSD
42 27 64.3 115 1 TVC2_MOUSE
43 27 64.3 117 1 RNPA_LACLA
44 27 64.3 132 1 TVC3_MOUSE
45 27 64.3 135 1 TVC1_MOUSE

```
CC -----
CC TRYPsin FAMILY. KALLIKREIN SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: L33404; AAC37551.1; -
CC EMBL: AF166330; AD49718.1; -
CC EMBL: AF243527; AAC33360.1; -
CC HSP: P00763; IDPO.
CC MEROPS: S01.300; -.
CC MIM: 604438; -.
CC InterPro: IPR001314; Chymotrypsin.
CC InterPro: IPR001254; Trypsin.
CC Pfam: PF00089; trypsin; 1.
CC PRINTS: PR00722; CHYMOTRYPSIN.
CC SMART: SM00020; Tryp_SPC; 1.
CC PROSITE: PS02040; TRYPSIN_DOM; 1.
CC PROSITE: PS00134; TRYPSIN_HIS; 1.
CC PROSITE: PS00135; TRYPSIN_SER; 1.
CC Hydrolase; Serine protease; Zymogen; Glycoprotein; Signal.
CC -----
CC FT SIGNAL 1 22
CC FT PROPEP 23 29
CC FT CHAIN 30 253
CC FT ACT_SITE 70 70
CC FT ACT_SITE 112 112
CC FT ACT_SITE 205 205
CC FT DISULFID 36 137
CC FT DISULFID 55 71
CC FT DISULFID 137 239
CC FT DISULFID 144 211
CC FT DISULFID 176 190
CC FT DISULFID 201 226
CC FT CARBOHYD 246 246
CC FT SEQUENCE 253 AA; 27525 MW; 2D69B6B15A76A668 CRC64;
CC -----
CC Query Match 100.0%; Score 42; DB 1; Length 253;
CC Best Local Similarity 100.0%; Pred. No. 0.036;
CC Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 1 QRIKAKSF 9
CC DB 91 QRIKAKSF 99
CC -----
CC RESULT 2
CC RLAO_THEAC
CC ID RLAO_THEAC STANDARD; PRT; 314 AA.
CC AC P57692;
CC DT 16-OCT-2001 (Rel. 40, Created)
CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Acidic ribosomal protein P0 homolog (L10E).
CC GN RPLP0 OR TA0359.
CC OS Thermoplasma acidophilum.
CC OC Archaea; Euryarchaeota; Thermoplasmales; Thermoplasmaceae;
CC OC Thermoplasma.
CC OX NCBI_TaxID=2303;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RC STRAIN=DSM 1728;
CC RX MEDLINE=20479972; PubMed=11029001;
CC RA Ruepp A., Graml W., Santos-Martinez M.-L., Koretke K.K., Volker C.,
CC RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.;
CC RT "The genome sequence of the thermoacidophilic scavenger Thermoplasma
CC RT acidophilum."
CC RL Nature 407:508-513(2000)
CC CC -1- FUNCTION: RIBOSOMAL PROTEIN P0 IS THE FUNCTIONAL EQUIVALENT
CC OF E.COLI PROTEIN L10.
CC -----
```

```
CC -----
CC -1- SIMILARITY: BELONGS TO THE L10P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL445064; CAC11503.1; -
CC InterPro: IPR001790; Ribosomal_L10.
CC Pfam: PF00466; Ribosomal_L10; 1.
CC KW Ribosomal protein; Complete proteome.
CC SQ SEQUENCE 314 AA; 34583 MW; F4D0D48680E0B6EB CRC64;
CC -----
CC Query Match 78.6%; Score 33; DB 1; Length 314;
CC Best Local Similarity 87.5%; Pred. No. 5;
CC Matches 7; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
CC -----
CC QY 1 QRIKAKS 8
CC DB 17 QRIKAKS 24
CC -----
CC RESULT 3
CC KAIN_HUMAN
CC ID KAIN_HUMAN STANDARD; PRT; 427 AA.
CC AC P29622;
CC DT 01-APR-1993 (Rel. 25, Created)
CC DT 01-OCT-1996 (Rel. 34, Last sequence update)
CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
CC DE Kallistatin precursor (Kallikrein inhibitor) (Protease inhibitor 4).
CC GN SERPIN4 OR PI4 OR KST.
CC OS Homo sapiens (Human).
CC OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
CC OC Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
CC OX NCBI_TaxID=9606;
CC RN [1]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=94043294; PubMed=8227002;
CC RA Chai K.X., Chen L.-M., Chao J., Chao L.;
CC RT "Kallistatin: a novel human serine proteinase inhibitor. Molecular
CC RT cloning, tissue distribution, and expression in Escherichia coli."
CC RL J. Biol. Chem. 268:24498-24505(1993).
CC RN [2]
CC RP SEQUENCE FROM N.A.
CC RX MEDLINE=95137583; PubMed=7835886;
CC RA Chai K.X., Ward D.C., Chao J., Chao L.;
CC RT "Molecular cloning, sequence analysis, and chromosomal localization
CC RT of the human protease inhibitor 4 (kallistatin) gene (PI4).";
CC RL Genomics 23:370-378(1994).
CC RN [3]
CC RP SEQUENCE OF 388-403.
CC RC TISSUE=Plasma;
CC RX MEDLINE=93100304; PubMed=1334488;
CC RA Zhou G.X., Chao L., Chao J.;
CC RT "Kallistatin: a novel human tissue kallikrein inhibitor.
CC RT Purification, characterization, and reactive center sequence."
CC RL J. Biol. Chem. 267:25873-25880(1992).
CC CC -1- FUNCTION: INHIBITS HUMAN AMIDOLYTIC AND KININOGENASE ACTIVITIES OF
CC HUMAN TISSUE KALLIKREIN. INHIBITION IS ACHIEVED BY FORMATION OF AN
CC EQUI-MOLAR, HEAT- AND SDS-STABLE COMPLEX BETWEEN THE INHIBITOR AND
CC THE ENZYME, AND GENERATION OF A SMALL C-TERMINAL FRAGMENT OF THE
CC INHIBITOR DUE TO CLEAVAGE AT THE REACTIVE SITE BY TISSUE
CC KALLIKREIN.
CC CC -1- SUBUNIT: MONOMER AND SOME HOMODIMERS.
CC CC -1- TISSUE SPECIFICITY: SECRETED FROM LIVER CELLS LINES.
CC CC -1- PTM: THE N-TERMINUS IS BLOCKED.
CC CC -1- MISCELLANEOUS: HEPARIN BLOCKS KALLISTATIN'S COMPLEX FORMATION WITH
CC TISSUE KALLIKREIN AND ABOLISHES ITS INHIBITORY EFFECT ON TISSUE
CC KALLIKREIN'S ACTIVITY.
CC CC -1- SIMILARITY: BELONGS TO THE SERPIN FAMILY.
CC -----
```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L19684; AA59454.1; -
CC EMBL; L28101; AAC41706.1; -
CC HSP; P05154; IPAI.
CC MIM; 147935; -
CC InterPro: IPR000215; Serpin.
CC Pfam; PF00079; serpin; 1.
CC SMART; SM00093; SERPIN; 1.
CC WSQSITE; PS00284; SERPIN; 1.
KW Serpin; Serine protease inhibitor; Plasma; Glycoprotein; Signal.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 427 KALLISTATIN.
FT ACT_SITE 389 389 REACTIVE BOND.
FT CARBOHYD 33 33 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 108 108 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 427 AA; 48556 MW; 3DBBE7AF956D4DAC CRC64;

Query Match 78.6%; Score 33; DB 1; Length 427;
Best Local Similarity 66.7%; Pred. NO. 6.9;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
   |:::|||||
Db 355 QKLEASKSF 363

RESULT 4
PCPC_FLAS3
ID PCPC_FLAS3 STANDARD; PRT; 247 AA.
AC Q03520;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Tetrachloro-p-hydroquinone reductive dehalogenase (BC 2.5.1.-).
GN PCPC.
OS Flavobacterium sp. (strain ATCC 39723).
OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonadaceae;
OC Sphingobium.
OX NCBI_TaxID=46429;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93239690; PubMed=8478329;
RA Orser C.S., Dutton J., Lange C.C., Jablonski P.E., Xun L., Hargis M.;
RT "Characterization of a Flavobacterium glutathione S-transferase gene
RT involved reductive dechlorination.";
RL J. Bacteriol. 175:2640-2644(1993).
RN [2]
RP SEQUENCE OF 1-18, AND CHARACTERIZATION.
RX MEDLINE=93094130; PubMed=1459949;
RA Xun L., Topp E., Orser C.S.;
RT "Purification and characterization of a tetrachloro-p-hydroquinone
RT reductive dehalogenase from a Flavobacterium sp.";
RL J. Bacteriol. 174:8003-8007(1992).
CC -1- FUNCTION: SEQUENTIAL REDUCTION OF TETRACHLORO-P-HYDROQUINONE TO
CC MONOCHLOROPHENOL, USING GLUTATHIONE AS THE REDUCING AGENT.
CC -1- PATHWAY: SECOND, THIRD AND FOURTH STEPS IN PENTACHLOROPHENOL (PCP)
CC DEGRADATION.
CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE GST SUPERFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
```

```
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M98559; AAA24921.1; -
CC PIR; A40625; A40625.
CC InterPro: IPR004046; GST_C.
CC InterPro: IPR004045; GST_N.
CC Pfam; PF00043; GST_C; 1.
CC Pfam; PF02798; GST_N; 1.
KW Aromatic hydrocarbons catabolism; Transferase.
FT INIT_MET 0
FT SEQUENCE 247 AA; 28115 MW; 263587F04BEC8F2B CRC64;

Query Match 73.8%; Score 31; DB 1; Length 247;
Best Local Similarity 66.7%; Pred. NO. 11;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
   |:::|||||
Db 226 QRMKARRSF 234

RESULT 5
SNAB_BOVIN STANDARD; PRT; 298 AA.
ID SNAB_BOVIN STANDARD; PRT; 298 AA.
AC P81126;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Beta-soluble NSF attachment protein (SNAP-beta) (N-ethylmaleimide-
DE sensitive factor attachment protein, beta).
GN NABP OR SNAPB.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC TISSUE=Brain;
RX MEDLINE=93205122; PubMed=8455721;
RA Whiteheart S.W., Griff I.C., Brunner M., Clary D.O., Mayer T.,
RA Buhrow S.A., Rothman J.E.;
RT "SNAP family of NSF attachment proteins includes a brain-specific
RT isoform.";
RL Nature 362:353-355(1993).
CC -1- FUNCTION: REQUIRED FOR VESICULAR TRANSPORT BETWEEN THE ENDOPLASMIC
CC RETICULUM AND THE GOLGI APPARATUS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC PERIPHERAL MEMBRANE PROTEIN (BY
CC similarity).
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: BELONGS TO THE SNAP FAMILY.
CC InterPro: IPR000744; NSF_attach.
CC Pfam; PF02071; NSF; 1.
CC PRINTS; PR00448; NSFATTACHMNT.
KW Transport; Protein transport; Endoplasmic reticulum; Golgi stack.
SQ SEQUENCE 298 AA; 33555 MW; CDA8912B93A6C27E CRC64;

Query Match 73.8%; Score 31; DB 1; Length 298;
Best Local Similarity 66.7%; Pred. NO. 13;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
   |:::|||||
Db 19 KRVKASHSF 27

RESULT 6
SNAB_HUMAN STANDARD; PRT; 298 AA.
ID SNAB_HUMAN STANDARD; PRT; 298 AA.
AC Q9H115;
```

16-OCT-2001 (Rel. 40, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 01-MAR-2002 (Rel. 41, Last annotation update)
 Beta-soluble NSF attachment protein (SNAP-beta) (N-ethylmaleimide-sensitive factor attachment protein, beta).
 NABP OR SNAPB.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A. (ISOFORM 1).
 Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R., Jones M., Stavrides G., Almeida J.P., Babbage A.K., Bagguley C.L., Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J., Buck D., Burrill W., Butler A.P., Carder C., Carter N.P., Clee C.M., Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clegg S., Clegg S., Cobley V.E., Collier R.E., Connor R., Corby N.R., Coulson A., Coville G.J., Deadman R., Dharm P., Dunn M., Ellington A.G., Griffiths J.A., Fraser A., French L., Garner P., Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J., Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D., Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S., Levasallano M.H., Leversha M., Lloyd C., Lloyd D.M., Lovell J.D., Marsh V.L., Martin S.L., McConnachie L.J., McLay K., McMurray A.A., Milne S., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T., Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I., Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H., Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S., Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E., Swann M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A., Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Whitehead S.L., Whittaker P., Willey D.L., Williams L., Williams S.A., Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S., Rogers J.;
 "The DNA sequence and comparative analysis of human chromosome 20.";
 Nature 414:865-871(2001).
 -1- FUNCTION: REQUIRED FOR VESICULAR TRANSPORT BETWEEN THE ENDOPLASMIC RETICULUM AND THE GOLGI APPARATUS (BY SIMILARITY).
 -1- SUBCELLULAR LOCATION: CYTOPLASMIC PERIPHERAL MEMBRANE PROTEIN (BY SIMILARITY).
 -1- SIMILARITY: BELONGS TO THE SNAP FAMILY.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).

 EMBL: AL096677; CAC03439.2; ALT_SEQ.
 InterPro: IPR000744; NSF_attach.
 Pfam: PF02071; NSF; 6.
 PRINTS: PR00448; NSFATACHMT.
 KW Transport; Protein transport; Endoplasmic reticulum; Golgi stack.
 SQ SEQUENCE 298 AA; 33557 MW; 5B7BE0FB84B8D83 CRC64;
 Query Match 73.8%; Score 31; DB 1; Length 298;
 Best Local Similarity 66.7%; Pred. No. 13;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 Qy 1 QIRKASKSF 9
 Db 19 KRVKASHSF 27
 :|:|:|:|:|
 RESULT 7
 TRIO_HUMAN STANDARD; PRT; 3038 AA.
 ID TRIO_HUMAN
 AC Q75962; Q13458;

16-OCT-2001 (Rel. 40, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 01-MAR-2002 (Rel. 41, Last annotation update)
 Triple functional domain protein (PtpRF interacting protein).
 TRIO.
 Homo sapiens (Human).
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TaxID=9606;
 [1]
 SEQUENCE FROM N.A. (ISOFORM 2).
 TISSUE=Fibroblast;
 MEDLINE=96224308; PubMed=8643598;
 Debat A., Serra-Pages C., Selpel K., O'Brien S., Tang M., Park S.-H., Streuli M.;
 "The multidomain protein Trio binds the LAR transmembrane tyrosine phosphatase, contains a protein kinase domain, and has separate rac-specific and rho-specific guanine nucleotide exchange factor domains.";
 Proc. Natl. Acad. Sci. U.S.A. 93:5466-5471(1996).
 [2]
 SEQUENCE FROM N.A. (ISOFORM 1).
 Streuli M.;
 Submitted (SEP-1998) to the EMBL/GenBank/DBJ databases.
 [3]
 STRUCTURE BY NMR OF 1227-1407, AND MUTAGENESIS.
 MEDLINE=99005194; PubMed=9790533;
 Liu X., Wang H., Eberstadt M., Schnuchel A., Olejniczak E.T., Meadows R.P., Schkeryantz J.M., Janowick D.A., Harlan J.E., Harris E.A.S., Staunton D.E., Fesik S.W.;
 "NMR structure and mutagenesis of the N-terminal Dbl homology domain of the nucleotide exchange factor Trio.";
 Cell 95:269-277(1998).
 [4]
 CHARACTERIZATION.
 MEDLINE=99276567; PubMed=10341202;
 Selpel K., Medley Q.G., Kedarsha N.L., Zhang X.A., O'Brien S.P., Serra-Pages C., Hemler M.E., Streuli M.;
 "Trio amino-terminal guanine nucleotide exchange factor domain expression promotes actin cytoskeleton reorganization, cell migration and anchorage-independent cell growth.";
 J. Cell Sci. 112:1825-1834(1999).
 -1- FUNCTION: PROMOTES THE EXCHANGE OF GDP BY GTP. TOGETHER WITH LEUCOCYTE ANTIGEN-RELATED (LAR) PROTEIN, IT COULD PLAY A ROLE IN COORDINATING CELL-MATRIX AND CYTOSKELETAL REARRANGEMENTS NECESSARY FOR CELL MIGRATION AND CELL GROWTH.
 -1- SUBUNIT: INTERACT TO FORM A COMPLEX WITH LEUCOCYTE ANTIGEN RELATED PROTEIN.
 -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE PRODUCED BY ALTERNATIVE SPLICING.
 -1- TISSUE SPECIFICITY: HIGHLY EXPRESSED IN HEART, SKELETAL MUSCLE, BRAIN, PANCREAS, PLACENTA, LIVER, KIDNEY AND LUNG.
 -1- DOMAIN: THE N-TERMINAL DBL/GEF DOMAIN SPECIFICALLY CATALYZES NUCLEOTIDE EXCHANGE FOR RAC1, LEADING TO THE ACTIVATION OF JUN KINASE AND THE PRODUCTION OF MEMBRANE RUFFLES. THE SECOND DBL/GEF DOMAIN IS AN EXCHANGE FACTOR FOR RHOA AND INDUCES THE FORMATION OF STRESS FIBERS.
 -1- PTM: SERINE PHOSPHORYLATED.
 -1- SIMILARITY: BELONGS TO THE RHO/RAC GEF FAMILY.
 -1- SIMILARITY: CONTAINS 2 DBL-HOMOLOGY DOMAINS (DH).
 -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE C2-TYPE DOMAIN.
 -1- SIMILARITY: CONTAINS 2 PH DOMAINS.
 -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 -1- SIMILARITY: CONTAINS 4 SPECTRIN REPEATS.
 -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/>).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z38113; CA86247.1; -;
 DR EMBL: X54237; CA38143.1; -;
 DR EMBL: X79743; CAB38096.1; -;
 DR PIR: C39610; C39610.
 DR SGD: S0001266; BET1.
 DR SMART: SM00397; tSNARE: 1.
 DR InterPro: IPR000727; tSNARE: 1.
 KW Transport; Protein transport; Golgi stack; Endoplasmic reticulum;
 KW Transmembrane.
 FT DOMAIN 1 117 CYTOPLASMIC (PROBABLE).
 FT TRANSMEM 118 141 MEMBRANE ANCHOR (BY SIMILARITY).
 SQ SEQUENCE 142 AA; 15720 MW; DC1BD4FFE67E2217 CRC64;

Query Match 71.4%; Score 30; DB 1; Length 142;
 Best Local Similarity 87.5%; Pred. No. 10;
 Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 1 QRIKASKS 8
 ||||| ||
 Db 67 QRIKALKS 74

RESULT 9
 YB3C_SCHPO STANDARD; PRT; 517 AA.
 ID YB3C_SCHPO
 AC O14342;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical 60.9 kDa protein C2F12.12C in chromosome II.
 GN SPBC2P12.12C.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Wood V., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;
 RL Submitted (JUN-1997) to the EMBL/GenBank/DBJ databases.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: Z97211; CAB10159.1; -;
 DR Hypothetical protein.
 KW SEQUENCE 517 AA; 60861 MW; 57C7C6B1E3B7F1E5 CRC64;

Query Match 71.4%; Score 30; DB 1; Length 517;
 Best Local Similarity 77.8%; Pred. No. 39;
 Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Oy 1 QRIKASKSF 9
 ||| | |||
 Db 187 QRIKAPKSF 195

RESULT 10
 PEPO_LACLA

ID PEPO_LACLA STANDARD; PRT; 626 AA.
 AC Q07744;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Neutral endopeptidase (EC 3.4.24.-) (Endopeptidase O).
 GN PEPO OR LL1803.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus.
 OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SSL135;
 RX MEDLINE=94064545; PubMed=8244921;
 RA Tynkkynen S., Buist G., Kunjl E., Kok J., Poolman B., Venema G.,
 RA Haandrikman A.;
 RT "Genetic and biochemical characterization of the oligopeptide
 RT transport system of Lactococcus lactis.";
 RL J. Bacteriol. 175:7523-7532(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ILI403;
 RX MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Wincker P., Mauger S., Jaillon O., Malarme K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 RT lactis ssp. lactis ILI403.";
 RL Genome Res. 11:731-753(2001).
 CC -1- FUNCTION: ENDOPEPTIDASE WITH BROAD SUBSTRATE SPECIFICITY FOR
 CC SEVERAL OLIGOPEPTIDES. OPTIMUM ACTIVITY IS MEASURED AT
 CC PH 6.0-6.5 AND TEMPERATURES 30-38 DEGREES C.
 CC -1- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -1- SUBUNIT: MONOMER.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M13 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE NEPRILYSIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: L18760; AAA16168.1; -;
 DR EMBL: AE006410; AAK05901.1; -;
 DR MEROPS; M13.004; -;
 DR InterPro: IPR000718; Peptidase_M13.
 DR InterPro: IPR000130; Zn_Mtpeptdse.
 DR Pfam: PF01431; Peptidase_M13; 1.
 DR PRINTS: PR00786; NEPRILYSIN
 DR PROSITE: PS00142; ZINC_PROTEASE; 1.
 KW Hydrolyase; Metalloprotease; Zinc; Complete proteome.
 FT INIT_MET 0
 FT METAL 474 474 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 475 475 BY SIMILARITY.
 FT METAL 478 478 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 534 534 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 538 538 PROTON DONOR (BY SIMILARITY).
 FT CONFLICT 75 75 T -> A (IN REF. 1).
 FT CONFLICT 216 216 S -> N (IN REF. 1).
 FT CONFLICT 350 350 A -> V (IN REF. 1).
 FT CONFLICT 406 406 E -> K (IN REF. 1).
 FT CONFLICT 470 470 T -> A (IN REF. 1).
 SQ SEQUENCE 626 AA; 71406 MW; D89FB627D25F59FD CRC64;

Query Match 71.4%; Score 30; DB 1; Length 626;
 Best Local Similarity 75.0%; Pred. No. 48;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Oy 2 RIRAKSKSF 9
 |::||| |

Db 569 RMKASREF 576

RESULT 11

PEPO_LACLC STANDARD; PRT; 626 AA.
 ID PEPO_LACLC
 AC Q09145;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Neutral endopeptidase (EC 3.4.24.-) (Endopeptidase O).
 GN PEPO.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Lactococcus.
 OX NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-20.
 RC STRAIN=P8-2-47;
 RX MEDLINE=93209954; PubMed=8458851;
 RA Mierau I., Tan P.S.T., Haandrikman A.J., Kok J., Leenhouts K.J.,
 RA Konings W.N., Venema G.;
 RT "Cloning and sequencing of the gene for a lactococcal endopeptidase,
 RT an enzyme with sequence similarity to mammalian enkephalinase.";
 RL J. Bacteriol. 175:2087-2096(1993).
 RN [2]
 RP SEQUENCE OF 1-286 FROM N.A.
 RC STRAIN=SK11;
 RA Yu W., Gillies K., Kondo J.K., Broadbent J.R., McKay L.L.;
 RL Submitted (MAY-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: ENDOPEPTIDASE WITH BROAD SUBSTRATE SPECIFICITY TO
 CC SEVERAL OLIGOPEPTIDES. OPTIMUM ACTIVITY IS MEASURED AT
 CC PH 6.0-6.5 AND TEMPERATURES 30-38 DEGREES C.
 CC -!- COFACTOR: BINDS 1 ZINC ION (BY SIMILARITY).
 CC -!- SUBUNIT: MONOMER.
 CC -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M13 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE NEPRILYSIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L04938; AAA35204.1; -;
 CC EMBL; U09553; AAB00538.1; -;
 CC MEROPS; M13.004; -;
 CC InterPro; IPR000718; Peptidase_M13.
 CC Pfam; PF01431; Peptidase_M13; 1.
 CC PRINTS; PR00786; NEPRILYSIN.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC KW Hydrolase; Metalloprotease; Zinc.
 CC FT INIT_MET 0
 CC FT METAL 474 474 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT ACT_SITE 475 475 BY SIMILARITY.
 CC FT METAL 478 478 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT METAL 534 534 ZINC (CATALYTIC) (BY SIMILARITY).
 CC FT ACT_SITE 538 538 PROTON DONOR (BY SIMILARITY).
 CC SEQUENCE 626 AA; 71396 MW; DDFED4FEE47FEE0C CRC64;

Query Match 71.4%; Score 30; DB 1; Length 626;
 Best Local Similarity 75.0%; Pred. No. 48;
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKASFSF 9

|:|||||

Db 569 RMKASREF 576

RESULT 12

PKP1_BOVIN STANDARD; PRT; 727 AA.
 ID PKP1_BOVIN
 AC Q28161;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DT 01-MAR-2002 (Rel. 41, Last annotation update)
 DE Plakophilin 1 (Band-6-protein) (B6P).
 GN PKP1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 OX NCBI_TaxID=9913;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95196971; PubMed=7890138;
 RA Heid H.W., Schmidt A., Zimbelmann R., Schaefer S.,
 RA Winter-Simonski S., Stupp S., Keith M., Figge U., Schnolzer M.,
 RA Franke W.W.;
 RT "Cell type-specific desmosomal plaque proteins of the plakoglobin
 RT family; plakophilin 1 (band 6 protein).";
 RL Differentiation 58:113-131(1994).
 CC -!- FUNCTION: SEEMS TO PLAY A ROLE IN JUNCTIONAL PLAQUES (BY
 CC SIMILARITY).
 CC -!- SUBCELLULAR LOCATION: Nuclear and associated with desmosomes (By
 CC similarity).
 CC -!- SIMILARITY: BELONGS TO THE BETA-CATENIN FAMILY.
 CC -!- SIMILARITY: CONTAINS 8 ARM REPEATS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; Z37975; CAA86029.1; -;
 CC InterPro; IPR002225; Armadillo.
 CC Pfam; PF00514; Armadillo_seg; 5.
 CC SMART; SM00185; ARM; 4.
 CC PROSITE; PS0176; ARM_REPEAT; 3.
 CC KW Cell adhesion; Cytoskeleton; Structural protein; Nuclear protein;
 CC Repeat.
 CC FT REPEAT 235 275 ARM 1.
 CC FT REPEAT 278 317 ARM 2.
 CC FT REPEAT 320 360 ARM 3.
 CC FT REPEAT 419 464 ARM 4.
 CC FT REPEAT 517 557 ARM 5.
 CC FT REPEAT 565 604 ARM 6.
 CC FT REPEAT 606 650 ARM 7.
 CC FT REPEAT 653 693 ARM 8.
 CC SEQUENCE 727 AA; 80180 MW; 3A27979279BCCEDF CRC64;
 Query Match 71.4%; Score 30; DB 1; Length 727;
 Best Local Similarity 75.0%; Pred. No. 56;
 Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
 |:|||||
 Db 149 QKIKASRS 156

RESULT 13
 PKP1_MOUSE STANDARD; PRT; 728 AA.
 ID PKP1_MOUSE
 AC P97350;
 DT 01-MAR-2002 (Rel. 41, Created)
 DT 01-MAR-2002 (Rel. 41, Last sequence update)
 DE Plakophilin 1.
 GN PKP1.
 OS Mus musculus (Mouse).

DT 01-OCT-1989 (Rel. 12, Last sequence update)
 DE 01-MAR-2002 (Rel. 41, Last annotation update)
 DN Dystrophin.
 GN DMD.
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89098331; PubMed=3062582;
 RA Lemaire C., Heilig R., Mandel J.L.;
 RT "Nucleotide sequence of chicken dystrophin cDNA."
 RL Nucleic Acids Res. 16:11815-11815(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Muscle;
 RX MEDLINE=89210800; PubMed=3072195;
 RA Lemaire C., Heilig R., Mandel J.L.;
 RT "The chicken dystrophin cDNA: striking conservation of the C-terminal
 coding and 3' untranslated regions between man and chicken."
 RL EMBO J. 7:4157-4162(1988).
 CC -!- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
 CC PLASMA MEMBRANE.
 CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
 CC ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
 CC ABP-120, ABP-180, OR BETA-TODRIN).
 CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
 CC -!- SIMILARITY: CONTAINS 22 SPECTRIN REPEATS.
 CC -!- SIMILARITY: CONTAINS 1 WW DOMAIN.
 CC -!- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X13369; CAA31746.1; -
 DR PIR: S02041; S02041.
 DR HSSP: P46939; 1BHD.
 DR InterPro: IPR001589; Actinin_act_bind.
 DR InterPro: IPR001715; Calponin_hom.
 DR InterPro: IPR002017; Spectrin.
 DR InterPro: IPR001202; WW.
 DR InterPro: IPR002349; WW_domain.
 DR InterPro: IPR000433; ZnF_Z2.
 DR Pfam: PF00307; CH; 2.
 DR Pfam: PF00435; spectrin; 22.
 DR Pfam: PF00397; WW; 1.
 DR Pfam: PF00369; Z2; 1.
 DR PRINTS: PR00403; WWDOMAIN.
 DR SMART: SM00033; CH; 2.
 DR SMART: SM00150; SPEC; 21.
 DR SMART: SM00456; WW; 1.
 DR SMART: SM00291; ZnF_Z2; 1.
 DR PROSITE: PS00019; ACTININ_1; 1.
 DR PROSITE: PS00020; ACTININ_2; 1.
 DR PROSITE: PS50021; CH; 2.
 DR PROSITE: PS01159; WW_DOMAIN_1; 1.
 DR PROSITE: PS50020; WW_DOMAIN_2; 1.
 DR PROSITE: PS01357; ZF_Z2_1; 2.
 DR PROSITE: PS50135; ZF_Z2_2; 1.
 DR Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
 KW Repeat; Zinc-finger.
 FT DOMAIN 1 244 ACTIN-BINDING.
 FT DOMAIN 19 123 CH 1.
 FT DOMAIN 138 241 CH 2.
 FT REPEAT 341 449 SPECTRIN 1.
 FT REPEAT 450 558 SPECTRIN 2.

FT REPEAT 561 669 SPECTRIN 3.
 FT REPEAT 721 830 SPECTRIN 4.
 FT REPEAT 832 936 SPECTRIN 5.
 FT REPEAT 945 1047 SPECTRIN 6.
 FT REPEAT 1050 1156 SPECTRIN 7.
 FT REPEAT 1159 1265 SPECTRIN 8.
 FT REPEAT 1268 1369 SPECTRIN 9.
 FT REPEAT 1470 1570 SPECTRIN 10.
 FT REPEAT 1573 1678 SPECTRIN 11.
 FT REPEAT 1681 1782 SPECTRIN 12.
 FT REPEAT 1879 1981 SPECTRIN 13.
 FT REPEAT 2013 2103 SPECTRIN 14.
 FT REPEAT 2106 2211 SPECTRIN 15.
 FT REPEAT 2214 2321 SPECTRIN 16.
 FT REPEAT 2472 2574 SPECTRIN 17.
 FT REPEAT 2577 2683 SPECTRIN 18.
 FT REPEAT 2686 2799 SPECTRIN 19.
 FT REPEAT 2802 2904 SPECTRIN 20.
 FT REPEAT 2906 2928 SPECTRIN 21.
 FT REPEAT 2931 3037 SPECTRIN 22.
 FT DOMAIN 3052 3085 WW.
 FT ZN_FING 3304 3351 ZZ-TYPE.
 FT VARIANT 1171 1171 MISSING.
 FT VARIANT 1869 1869 Q -> H.
 FT VARIANT 1885 1885 K -> R.
 SQ SEQUENCE 3660 AA; 422874 MW; 85493DAF6D5B6D4A CRC64;

 Query Match 71.4%; Score 30; DB 1; Length 3660;
 Best Local Similarity 66.7%; Pred. No. 3.1e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

 QY 1 QRIKASKSF 9
 Db 1771 QRIKASKPF 1779
 ||||| :|
 V16K TRVPL STANDARD; PRT; 141 AA.
 ID V16K TRVPL
 AC P33777;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 01-FEB-1994 (Rel. 28, Last sequence update)
 DT 01-FEB-1994 (Rel. 28, Last annotation update)
 DE 16 kDa protein.
 OS Tobacco rattle virus (strain PUB).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobravirus.
 OX NCBI_TaxID=33766;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89299469; PubMed=2741345;
 RA Angenent G.C., Posthumus E., Brederode F.T.M., Bol J.F.;
 RT "Genome structure of tobacco rattle virus strain PUB: further
 RT evidence on the occurrence of RNA recombination among tobnaviruses.";
 RL Virology 171:271-274(1989).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J04347; AAA47079.1; -
 DR PIR: B46322; B46322.
 SQ SEQUENCE 141 AA; 16331 MW; 4E4EFA194B8C6A20 CRC64;

 Query Match 69.0%; Score 29; DB 1; Length 141;
 Best Local Similarity 55.6%; Pred. No. 17;
 Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

 QY 1 QRIKASKSF 9
 ||||| :|

```
Db 90 ERIRAERSF 98

RESULT 17
OR_N_HAEIN STANDARD; PRT; 182 AA.
ID P45340;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Oligoribonuclease (EC 3.1.-.-).
GN ORN OR H11715
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kierlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd."
RL Science 269:496-512(1995).
CC -!- FUNCTION: 3'-TO-5' EXORIBONUCLEASE SPECIFIC FOR SMALL
CC OLIGORIBONUCLEOTIDES (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: BELONGS TO THE OLIGORIBONUCLEASE FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32844; AAC23360.1; -.
DR TIGR; H11715; -.
DR InterPro; IPR000520; Exonuclease.
DR Pfam; PF00929; Exonuclease; 1.
DR SMART; SM00479; EXOIII; 1.
DR Hydrolase; Exonuclease; Nuclease; Complete proteome.
FT ACT_SITE 129 129 POTENTIAL.
SQ SEQUENCE 182 AA; 21199 MW; B78FF5D619915D29 CRC64;

Query Match 69.0%; Score 29; DB 1; Length 182;
Best Local Similarity 85.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASK 7
Db 73 ERIKASK 79
:|||||

RESULT 18
SYN_CLOAB STANDARD; PRT; 463 AA.
ID Q97E56;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Asparaginyl-tRNA synthetase (EC 6.1.1.22) (Asparagine--tRNA ligase)
DE (AsnRS).
GN ASNS OR CAC3260.

Query Match 69.0%; Score 29; DB 1; Length 182;
Best Local Similarity 85.7%; Pred. No. 22;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASK 7
Db 73 ERIKASK 79
:|||||

RESULT 19
6PGD_ACTAC STANDARD; PRT; 484 AA.
ID 6PGD_ACTAC
AC P70718;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE 6-phosphogluconate dehydrogenase, decarboxylating (EC 1.1.1.44).
GN GND.
OS Actinobacillus actinomycetemcomitans (Haemophilus
OS actinomycetemcomitans).
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Actinobacillus.
OX NCBI_TaxID=714;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Y4;
RX MEDLINE=97148607; PubMed=9020051;
RA Yoshida Y., Nakano Y., Yamashita Y., Koga T.;
RT "The gnd gene encoding a novel 6-phosphogluconate dehydrogenase and
RT its adjacent region of Actinobacillus actinomycetemcomitans
```

RT chromosomal DNA.";
RL Biochem. Biophys. Res. Commun. 230:220-225(1997).
CC -1- CATALYTIC ACTIVITY: 6-phospho-D-glucuronate + NADP(+) = D-ribose
CC 5-phosphate + CO(2) + NADPH.
CC -1- PATHWAY: HEXOSE MONOPHOSPHATE SHUNT.
CC -1- SIMILARITY: BELONGS TO THE 6-PHOSPHOGLUCONATE DEHYDROGENASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D88189; BAAJ3558.1; -.
CC HSSP: P00349; 2PGD.
DR InterPro: IPR001744; 6PGD.
DR Pfam: PF00393; 6PGD; 1.
DR PRINTS: PR00076; 6PGDHRGNASE.
DR PROSITE: PS00461; 6PGD; 1.
KW Glucuronate utilization; Oxidoreductase; Pentose shunt; NADP.
SQ SEQUENCE 484 AA; 53288 MW; F7BD0B2EA3BF624D CRC64;

Query Match 69.0%; Score 29; DB 1; Length 484;
Best Local Similarity 66.7%; Pred. No. 62;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 ORTKASKSF 9
II: III I
Db 297 ORVAASKLF 305

RESULT 20
CALI_HUMAN STANDARD; PRT; 588 AA.
ID Q13939; Q9BXG7;
AC Q13939; Q9BXG7;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Calicin.
DE CCIN.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RA Sha J.H., Zhou Z.M., Li J.M.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 4-588 FROM N.A.
RC TISSUE=Sperm;
RX MEDLINE=95369354; PubMed=7641791;
RA von Buelow M., Held H.W., Hess H., Franke W.W.;
RT "Molecular nature of calicin, a major basic protein of the mammalian
RL sperm head cytoskeleton.";
RL Exp. Cell Res. 219:407-413(1995).
CC -1- FUNCTION: POSSIBLE MORPHOGENIC CYTOSKELETAL ELEMENT IN
CC SPERMIOGENIC DIFFERENTIATION.
CC -1- SUBCELLULAR LOCATION: CALYX; SPERM HEAD CYTOSKELETAL STRUCTURE
CC TIGHTLY ASSOCIATED TO THE NUCLEUS.
CC -1- TISSUE SPECIFICITY: TESTIS.
CC -1- DISEASE: NOT DETECTABLE OR SHOWS A DRASTICALLY ALTERED PATTERN OF
CC ARRANGEMENT IN THE HEADS OF MALFORMED SPERMATOZOA SUCH AS IN
CC TERATOZOOSPERMIAS WITH 'ROUND-HEADED' SPERM OR WITH OTHER
CC 'POSTACROSOMAL SHEATH DEFECTS'.
CC -1- SIMILARITY: CONTAINS 1 BTB/POZ DOMAIN.
CC -1- SIMILARITY: CONTAINS 4 KELCH REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF333334; AAK20829.1; -.
CC MIM: 603960; -.
DR InterPro: IPR000210; BTB_POZ.
DR InterPro: IPR001798; Kelch.
DR Pfam: PF00651; BTB; 1.
DR Pfam: PF01344; Kelch; 4.
DR SMART: SM00225; BTB; 1.
DR PROSITE: PS50097; BTB; 1.
KW Cytoskeleton; Spermatogenesis; Repeat.
FT DOMAIN 12 124
FT REPEAT 280 327
FT REPEAT 328 375
FT REPEAT 377 423
FT REPEAT 526 580
FT REPEAT 526 580
FT CONFLICT 226 226 S -> N (IN REF. 2).
FT CONFLICT 433 433 C -> W (IN REF. 2).
FT CONFLICT 490 490 N -> I (IN REF. 2).
SQ SEQUENCE 588 AA; 66555 MW; BD8F695F0529E63E CRC64;

Query Match 69.0%; Score 29; DB 1; Length 588;
Best Local Similarity 62.5%; Pred. No. 76;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORTKASKS 8
II: IIII I
Db 493 QKVKASKT 500

RESULT 21
DIA3_HUMAN STANDARD; PRT; 853 AA.
ID DIA3_HUMAN STANDARD; PRT; 853 AA.
AC Q9NSV4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Diaphanous protein homolog 3 (Diaphanous-related formin 3) (DRF3)
DE (Fragment).
DE DIAH3 OR DIAP3.
GN Homo sapiens (Human).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Testis;
RA Blum H., Bauersachs S., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 468-853 FROM N.A. (ISOFORMS 1 AND 2).
RA Smith M.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: BINDS TO GTP-BOUND FORM OF RHO AND TO PROFILIN. ACTS IN
CC A RHO-DEPENDENT MANNER TO RECRUIT PROFILIN TO THE MEMBRANE, WHERE
CC IT PROMOTES ACTIN POLYMERIZATION. IT IS REQUIRED FOR CYTOKINESIS,
CC STRESS FIBER FORMATION, AND TRANSCRIPTIONAL ACTIVATION OF THE
CC SERUM RESPONSE FACTOR. DFR PROTEINS COUPLE RHO AND SRC TYROSINE
CC KINASE DURING SIGNALING AND THE REGULATION OF ACTIN DYNAMICS (BY
CC SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; 1 (SHOWN HERE) AND 2; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: DRFS ARE REGULATED BY INTRAMOLECULAR GBD-DAD BINDING WHERE
CC RHO-GTP ACTIVATES THE DRFS BY DISRUPTING THE GBD-DAD INTERACTION
CC (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 GTPASE-BINDING DOMAIN (GBD).
CC -1- SIMILARITY: CONTAINS 1 FORMIN HOMOLOG 1 (FHL) DOMAIN.

DB 34 EQIKANKDF 42

RESULT 24

CSRA_ECOLI STANDARD; PRT; 61 AA.

AC P31803;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Carbon storage regulator

GN CSRA OR ZFTA OR B2696 OR Z3998 OR ECS3553 OR STM2826 OR STY2947.

OS Escherichia coli,

OS Escherichia coli O157:H7.

OS Salmonella typhimurium, and

OS Salmonella typhi.

OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

OC Escherichia.

OX NCBI_TaxID=562, 83334, 602, 601;

[1]

RN SEQUENCE FROM N.A.

RP SPECIES=E.coli; STRAIN=K12;

RX MEDLINE=93328679; PubMed=89393005;

RA Romeo T., Gong M., Liu M.-Y., Brun-Zinkernagel A.-M.;

RT "Identification and molecular characterization of csrA, a pleiotropic

RT gene from Escherichia coli that affects glycogen biosynthesis,

RT gluconeogenesis, cell size, and surface properties.";

RT J. Bacteriol. 175:4744-4755(1993).

[2]

RN SEQUENCE FROM N.A.

RP SPECIES=E.coli; STRAIN=K12 / MG1655;

RX MEDLINE=97426617; PubMed=9278503;

RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,

RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,

RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,

RA Mau B., Shao Y.;

RT "The complete genome sequence of Escherichia coli K-12.";

RT Science 277:1453-1474(1997).

[3]

RN SEQUENCE FROM N.A.

RP SPECIES=E.coli; STRAIN=K12;

RX MEDLINE=97349980; PubMed=9205837;

RA Yamamoto Y., Aiba H., Baba T., Hayashi K., Inada T., Isono K.,

RA Itoh T., Kimura S., Kitagawa M., Makino K., Miki T., Mitsuhashi N.,

RA Mizobuchi K., Mori H., Nakade S., Nakamura Y., Nishimoto H.,

RA Tagami H., Takahashi H., Takeda J., Takemoto K., Uehara K., Wada C.,

RA Yamagata S., Horiuchi T.;

RT "Construction of a contiguous 874-kb sequence of the Escherichia coli

RT - K12 genome corresponding to 50.0-68.8 min on the linkage map and

RT analysis of its sequence features.";

RT DNA Res. 4:91-113(1997).

[4]

RN SEQUENCE FROM N.A., AND CHARACTERIZATION.

RP SPECIES=E.coli; STRAIN=KP4714;

RX MEDLINE=96177756; PubMed=8604133;

RA Murayama N., Shimizu H., Takiguchi S., Baba Y., Amino H., Horiuchi T.,

RA Sekimizu K., Miki T.;

RT "Evidence for involvement of Escherichia coli genes pmbA, csrA and a

RT previously unrecognized gene tldD, in the control of DNA gyrase by

RT letD (ccdB) of sex factor F.";

RT J. Mol. Biol. 256:483-502(1996).

[5]

RN SEQUENCE FROM N.A.

RP SPECIES=E.coli; STRAIN=O157:H7 / EDL933 / ATCC 700927;

RX MEDLINE=21074935; PubMed=11206551;

RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,

RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,

RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,

RA Grobbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,

RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,

RA Welch R.A., Blattner F.R.;

RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";

RT Nature 409:529-533(2001).

[6]

RN SEQUENCE FROM N.A.

RP SPECIES=E.coli; STRAIN=O157:H7 / RIMD 0509952;

RX MEDLINE=21156231; PubMed=11258796;

RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,

RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,

RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;

RT "Complete genome sequence of enterohaemorrhagic Escherichia coli

RT O157:H7 and genomic comparison with a laboratory strain K-12.";

RT DNA Res. 8:11-22(2001).

[7]

RN SEQUENCE FROM N.A.

RP SPECIES=S.typhimurium; STRAIN=SR-11;

RA White D.L., Romeo T.;

RT "Molecular cloning and characterization of the Salmonella typhimurium

RT csrA gene.";

RT Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.

[8]

RN SEQUENCE FROM N.A.

RP SPECIES=S.typhimurium; STRAIN=ATCC 140285;

RX MEDLINE=20536422; PubMed=11083797;

RA Altier C., Suyemoto M., Lawhon S.D.;

RT "Regulation of Salmonella enterica serovar typhimurium invasion genes

RT by csrA.";

RT Infect. Immun. 68:6790-6797(2000).

[9]

RN SEQUENCE FROM N.A.

RP SPECIES=S.typhimurium; STRAIN=LT2 / SGSC1412 / ATCC 700720;

RX MEDLINE=21534948; PubMed=11677609;

RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,

RA Courtney L., Porwollik S., All J., Dante M., Du F., Hou S., Layman D.,

RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,

RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,

RA Waterston R., Wilson R.K.;

RT "Complete genome sequence of Salmonella enterica serovar Typhimurium

RT LT2.";

RT Nature 413:852-856(2001).

[10]

RN SEQUENCE FROM N.A.

RP SPECIES=S.typhi; STRAIN=CT18;

RX MEDLINE=21534947; PubMed=11677608;

RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Wain J.,

RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Sebahia M.,

RA Baker S., Basham D., Brooks K., Chillingworth T., Connor P.,

RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,

RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,

RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,

RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,

RA Whitehead S., Barrell B.G.;

RT "Complete genome sequence of a multiple drug resistant Salmonella

RT enterica serovar Typhi CT18.";

RT Nature 413:848-852(2001).

[11]

RN CHARACTERIZATION.

RP SPECIES=E.coli;

RX MEDLINE=95270581; PubMed=7751274;

RA Liu M.Y., Yang H., Romeo T.;

RT "The product of the pleiotropic Escherichia coli gene csrA modulates

RT glycogen biosynthesis via effects on mRNA stability.";

RT J. Bacteriol. 177:2663-2672(1995).

[12]

RN INTERACTION WITH CSRB.

RP SPECIES=E.coli;

RX MEDLINE=97362239; PubMed=9211896;

RA Liu M.Y., Gui G., Wei B., Preston J.F. III, Oakford L., Yuksel U.,

RA Giedroc D.P., Romeo T.;

RT "The RNA molecule csrB binds to the global regulatory protein CsrA

RT and antagonizes its activity in Escherichia coli.";

RT J. Biol. Chem. 272:17502-17510(1997).

[13]

RN FUNCTION: AFFECTS GLYCOGEN BIOSYNTHESIS, GLUCONEOGENESIS, CELL

CC SIZE AND SURFACE PROPERTIES. REGULATES GLYCOGEN SYNTHESIS UNDER
 CC BOTH AEROBIC AND ANAEROBIC CONDITIONS. SEEMS TO ACCELERATE THE
 CC DEGRADATION OF GLG GENES TRANSCRIPTS POTENTIALLY THROUGH SELECTIVE
 CC RNA BINDING. ACTS TO INHIBIT INTERACTION BETWEEN THE LETD PROTEIN
 CC AND THE A SUBUNIT OF DNA GYRASE.
 CC
 CC -1- SUBUNIT: BINDS TO THE CSRB RNA WHICH ANTAGONIZES THE ACTIVITY OF
 CC CSRA.
 CC
 CC -1- SIMILARITY: BELONGS TO THE CSRA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L07596; AAA71919.1; -
 CC EMBL; AE000353; AAC75738.1; -
 CC EMBL; D90892; BAA16558.1; -
 CC EMBL; D4453; BAA21555.1; -
 CC EMBL; AE005498; AAG57800.1; -
 CC EMBL; AP002562; BAB36976.1; -
 CC EMBL; AF161596; ABP80413.1; -
 CC EMBL; AF203376; AAG35184.1; -
 CC EMBL; AE008829; AAL21706.1; -
 CC EMBL; AL627276; CAD05932.1; -
 CC PIR; B40608; B40608.
 CC EcoGene; EG11447; csra.
 CC StyGene; SG22222; csra.
 CC InterPro; IPR003751; Csra.
 CC Pfam; PF02599; Csra; 1.
 CC RNA-binding; Complete proteome.
 CC FT DOMAIN 11 45 KH.
 CC SEQUENCE 61 AA; 6856 MW; 16308FC572670E1C CRC64;
 CC
 CC Query Match 66.7%; Score 28; DB 1; Length 61;
 CC Best Local Similarity 75.0%; Pred. No. 12;
 CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 1 QRIKASKS 8
 CC III:|I|I|
 CC DB 49 QRIQAQKS 56
 CC
 CC RESULT 25
 CC CSRA_ERWCA
 CC ID CSRA_ERWCA STANDARD; PRT; 61 AA.
 CC AC Q47620; Q9XB50;
 CC DT 01-NOV-1997 (Rel. 35, Created)
 CC DT 01-NOV-1997 (Rel. 35, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Carbon storage regulator homolog (Repressor RSMA).
 CC CSRA OR RSMA.
 CC GN Erwinia carotovora.
 CC OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Pectobacterium.
 CC OX NCBI_TaxID=554;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=71;
 CC RX MEDLINE=95394817; PubMed=7665490;
 CC RA Cui Y., Chatterjee A., Liu Y., Dumenyo C.K., Chatterjee A.K.;
 CC "Identification of a global repressor gene, rsma, of Erwinia
 CC carotovora subsp. carotovora that controls extracellular enzymes,
 CC N-(3-oxohexanoyl)-L-homoserine lactone, and pathogenicity in
 CC soft-rotting Erwinia spp.;
 CC J. Bacteriol. 177:5108-5115(1995).
 CC [2]
 CC RN SEQUENCE FROM N.A.
 CC RP STRAIN=SCC3193;
 CC RC Andersson R.A.;
 CC RA Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: CONTROLS EXTRACELLULAR ENZYMES, N-(3-OXOHXANOL)-L-
 CC HOMOSERINE LACTONE, AND PATHOGENICITY. REPRESSOR OF VIRULENCE
 CC FACTORS.
 CC
 CC -1- SIMILARITY: BELONGS TO THE CSRA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; L40173; AAA74502.1; -
 CC EMBL; AJ238885; CAB46440.1; -
 CC InterPro; IPR003751; Csra.
 CC Pfam; PF02599; Csra; 1.
 CC RNA-binding.
 CC FT DOMAIN 11 45 KH.
 CC SEQUENCE 61 AA; 6821 MW; 112BE66502670E1C CRC64;
 CC
 CC Query Match 66.7%; Score 28; DB 1; Length 61;
 CC Best Local Similarity 75.0%; Pred. No. 12;
 CC Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
 CC
 CC QY 1 QRIKASKS 8
 CC III:|I|I|
 CC DB 49 QRIQAQKS 56
 CC
 CC RESULT 26
 CC CSRA_SERMA
 CC ID CSRA_SERMA STANDARD; PRT; 69 AA.
 CC AC Q85735;
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DT 16-OCT-2001 (Rel. 40, Last annotation update)
 CC DE Carbon storage regulator homolog.
 CC CSRA OR RSMA.
 CC GN Serratia marcescens.
 CC OS Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 CC Serratia.
 CC OX NCBI_TaxID=615;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RA Ang S., Horng Y.-T., Shu J.-C., Soo P.-C., Liu J.-H., Yi W.-C.,
 CC Lai H.-C., Luh K.-T., Ho S.-W., Swift S.;
 CC "A role for RsmA in the regulation of swarming motility in Serratia
 CC marcescens".
 CC RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases
 CC -1- FUNCTION: INVOLVED IN THE PROCESS OF SWARMING AND QUORUM-SENSING
 CC SIGNAL PRODUCTION. COULD ACCELERATE THE DEGRADATION OF SOME GENES
 CC TRANSCRIPTS POTENTIALLY THROUGH SELECTIVE RNA BINDING
 CC (BY SIMILARITY).
 CC
 CC -1- SIMILARITY: BELONGS TO THE CSRA FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AF074437; AAC25783.1; -
 CC EMBL; AJ243121; CAB45388.1; -
 CC InterPro; IPR003751; Csra.
 CC Pfam; PF02599; Csra; 1.
 CC RNA-binding.
 CC FT DOMAIN 11 45 KH.
 CC SEQUENCE 69 AA; 7696 MW; 2256A78C31BD3F50 CRC64;

Query Match 66.7%; Score 28; DB 1; Length 69;
Best Local Similarity 75.0%; Pred. No. 13;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
DB 49 QRIQAEKS 56

RESULT 27
GREA_BACSU STANDARD; PRT; 156 AA.
ID GREA_BACSU STANDARD; PRT; 156 AA.
AC P80240;
DT 01-FEB-1995 (Rel. 31, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transcription elongation factor greA (transcript cleavage factor
greA) (General stress protein 20M) (GSP20M).
GN GREA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-27.
RC STRAIN=IS58;
RX MEDLINE=97443988; PubMed=9298659;
RA Antelmann H., Bernhardt J., Schmid R., Mach H., Voelker U.,
RA Hecker M.;
RT "First steps from a two-dimensional protein index towards a response-
regulation map for Bacillus subtilis.";
RL Electrophoresis 18:1451-1463(1997).
RN [3]
RP PRELIMINARY SEQUENCE OF 1-17.
RC STRAIN=IS58;
RX MEDLINE=94282319; PubMed=8012595;
RA Voelker U., Engelmann S., Maul B., Riethdorf S., Voelker A.,
RA Schmid R., Mach H., Hecker M.;
RT "Analysis of the induction of general stress proteins of Bacillus
subtilis.";
RL Microbiology 140:741-752(1994).
CC -!- FUNCTION: NECESSARY FOR EFFICIENT RNA POLYMERASE TRANSCRIPTION
OF ELONGATION PAST TEMPLATE-ENCODED ARRESTING SITES. THE ARRESTING
SITES IN DNA HAVE THE PROPERTY OF TRAPPING A CERTAIN FRACTION
OF ELONGATING RNA POLYMERASES THAT PASS THROUGH, RESULTING IN
LOCKED TERNARY COMPLEXES. CLEAVAGE OF THE NASCENT TRANSCRIPT BY
CLEAVAGE FACTORS SUCH AS GREA OR GREB ALLOWS THE RESUMPTION OF
ELONGATION FROM THE NEW 3' TERMINUS. GREA RELEASES SEQUENCES OF
2 TO 3 NUCLEOTIDES (BY SIMILARITY).
CC -!- INDUCTION: BY HEAT SHOCK, SALT STRESS, OXIDATIVE STRESS, GLUCOSE
LIMITATION AND OXYGEN LIMITATION.
CC -!- SIMILARITY: BELONGS TO THE GREA/GREB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; Z99117; CAB14674.1; -
DR HSSP; P21346; IGRJ.
DR Subtilist; BG19019; greA.
DR InterPro; IPR001437; Grea_Greb.
DR Pfam; PF01272; Grea_Greb; 1.
DR ProDom; PD004918; Grea_Greb; 1.
DR PROSITE; PS00829; Grea_Greb; 1.
DR PROSITE; PS00829; Grea_Greb; 1.

DR PROSITE; PS00830; GREAB_2; 1.
KW Transcription regulation; DNA-binding; Coiled coil; Heat shock;
KW Complete proteome.
FT INIT_MET 0
FT DOMAIN 46 76 COILED COIL (POTENTIAL).
SQ SEQUENCE 156 AA; 17140 MW; CBCA4FFB3A108016 CRC64;

Query Match 66.7%; Score 28; DB 1; Length 156;
Best Local Similarity 55.6%; Pred. No. 32;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
DB 33 ERIKIARSF 41

RESULT 28
GREA_BACHD STANDARD; PRT; 158 AA.
ID GREA_BACHD STANDARD; PRT; 158 AA.
AC Q9KDD7;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Transcription elongation factor greA (Transcript cleavage factor
greA).
GN GREA OR BH1276.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125/JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
CC -!- FUNCTION: Necessary for efficient RNA polymerase transcription
elongation past template-encoded arresting sites. The arresting
sites in DNA have the property of trapping a certain fraction of
elongating RNA polymerases that pass through, resulting in locked
ternary complexes. Cleavage of the nascent transcript by cleavage
factors such as greA or greB allows the resumption of elongation
from the new 3' terminus. GreA releases sequences of 2 to 3
nucleotides (by similarity).
CC -!- SIMILARITY: BELONGS TO THE GREA/GREB FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).
CC
CC EMBL; AP001511; BAB04995.1; -
DR HSSP; P21346; IGRJ.
DR InterPro; IPR001437; Grea_Greb.
DR Pfam; PF01272; Grea_Greb; 1.
DR ProDom; PD004918; Grea_Greb; 1.
DR PROSITE; PS00829; Grea_Greb; 1.
DR PROSITE; PS00830; GREAB_2; 1.
KW Transcription regulation; DNA-binding; Coiled coil; Complete proteome.
FT DOMAIN 11 80 COILED COIL (POTENTIAL).
SQ SEQUENCE 158 AA; 17855 MW; 565CAEFD8FF8EA2 CRC64;

Query Match 66.7%; Score 28; DB 1; Length 158;
Best Local Similarity 55.6%; Pred. No. 32;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9
Db 34 ERIKARSF 42

```
RESULT 29
MK16_YEAST
ID MK16_YEAST STANDARD; PRT; 306 AA.
AC P10962;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-JUL-1989 (Rel. 11, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE MAK16 protein.
GN MAK16 OR YAL025C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=88320371; PubMed=3045810;
RA Wickner R.B.;
RT "Host function of MAK16: G1 arrest by a mak16 mutant of Saccharomyces cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 85:6007-6011(1988).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=S288C / AB972;
RX MEDLINE=95249563; PubMed=7731988;
RA Bussey H., Kaback D.B., Zhong W., Vo D.T., Clark M.W., Fortin N., Hall J., Ouellette B.F., Keng T., Barton A.B., Su Y., Davies C.K., Storms R.K.;
RT "The nucleotide sequence of chromosome I from Saccharomyces cerevisiae."
RL Proc. Natl. Acad. Sci. U.S.A. 92:3809-3813(1995).
CC -1- FUNCTION: ITS ROLE MIGHT BE AS PART OF THE APPARATUS CONCERNED WITH THE NUCLEAR EVENTS OF THE CELL CYCLE.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: NONE TO THE GST SUPERFAMILY.
CC -1- SIMILARITY: STRONG, TO S.MANSONI GLUTATHIONE S-TRANSFERASE 6 AND S.POMBE SPAC222.06.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
DR EMBL; J03852; AAA34752.1; -.
DR EMBL; U12980; AAC05007.1; -.
DR PIR; S05851; BVBYK6.
DR SGD; S0000023; MAK16.
KW Cell cycle; Nuclear protein; Repeat; Phosphorylation.
FT DOMAIN 139 144 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 195 235 ASP/GLU-RICH (ACIDIC).
FT DOMAIN 213 229 2 X 8 AA REPEATS.
FT REPEAT 213 220 1.
FT REPEAT 222 229 2.
FT DOMAIN 251 264 7 X 2 AA REPEATS OF S-[DE].
FT DOMAIN 282 287 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT MOD_RES 212 212 PHOSPHORYLATION (POTENTIAL).
FT MOD_RES 241 241 PHOSPHORYLATION (POTENTIAL).
SQ SEQUENCE 306 AA; 248AD69D6AA24F73 CRC64;

Query Match 66.7%; Score 28; DB 1; Length 306;
Best Local Similarity 55.6%; Pred. No. 64;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9
Db 78 ERIKSKNY 86
```

```
RESULT 30
MURB_BUCAI
ID MURB_BUCAI STANDARD; PRT; 356 AA.
AC P57153;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE UDP-N-acetylenolpyruvoylglucosamine reductase (EC 1.1.1.158) (UDP-N-acetylmuramate dehydrogenase).
GN MURB OR BU045.
OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum symbiotic bacterium).
OC Bacteria; Proteobacteria; gamma subdivision; Buchnera.
OX NCBI_TaxID=118099;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=TOKYO 1998;
RX MEDLINE=20445173; PubMed=10993077;
RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
RT "Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. AFS."
RL Nature 407:81-86(2000).
CC -1- FUNCTION: CELL WALL FORMATION (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: UDP-N-acetylmuramate + NADP(+) = UDP-N-acetyl-3-O-(1-carboxyvinyl)-D-glucosamine + NADPH.
CC -1- COFACTOR: FAD (BY SIMILARITY).
CC -1- PATHWAY: PEPTIDOGLYCAN BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: BELONGS TO THE MURB FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
DR EMBL; AP001118; BAB12768.1; -.
DR InterPro; IPR003170; MurB.
DR Pfam; PF02215; MurB; 1.
DR Pfam; PF02873; MurB_C; 1.
KW Peptidoglycan synthesis; Cell wall; Cell division; Oxidoreductase; NADP; Flavoprotein; FAD; Complete proteome.
SQ SEQUENCE 356 AA; 40592 MW; 103486BF1DE25EA7 CRC64;

Query Match 66.7%; Score 28; DB 1; Length 356;
Best Local Similarity 75.0%; Pred. No. 75;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKASKSF 9
Db 340 KIRASKIF 347

RESULT 31
PGK_YEAST
ID PGK_YEAST STANDARD; PRT; 415 AA.
AC P00560;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Phosphoglycerate kinase (EC 2.7.2.3).
GN PGK1 OR YCR012W OR YCR12W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=83116992; PubMed=6296791;
```


RA Hitzeman R.A., Hagie F.E., Hayflick J.S., Chen C.Y., Seeburg P.H.,
RA Derynck R.;
RT "The primary structure of the Saccharomyces cerevisiae gene for 3-
RT phosphoglycerate kinase.";
RL Nucleic Acids Res. 10:7791-7808(1982).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=92327849; PubMed=1626432;
RA Skala J., Purnelle B., Goffeau A.;
RT "The complete sequence of a 10.8 kb segment distal of SUP2 on the
RT right arm of chromosome III from Saccharomyces cerevisiae reveals
RT seven open reading frames including the RVS161, ADP1 and PGK genes.";
RL Yeast 4:409-417(1992).
RN [3]
RP SEQUENCE OF 172-201 AND 237-269.
RX MEDLINE=82029924; PubMed=7287307;
RA Fattoum A., Roustan C., Karoui D., Feinberg J., Pradel L.-A.,
RA Gregoire J., Rochat H.;
RT "Structural studies on yeast 3-phosphoglycerate kinase. Linear
RT arrangement of the CNBR fragments, partial amino acid sequence of the
RT inner part of the polypeptide chain, and analyses of the N-terminal
RT domain of the protein.";
RL Int. J. Pept. Protein Res. 17:393-400(1981).
RN [4]
RP SEQUENCE OF 281-415 FROM N.A.
RX MEDLINE=83256430; PubMed=6347186;
RA Perkins R.E., Conroy S.C., Dunbar B., Fothergill L.A., Tuite M.F.,
RA Dobson M.J., Kingsman S.M., Kingsman A.J.;
RT "The complete amino acid sequence of yeast phosphoglycerate kinase.";
RL Biochem. J. 211:199-218(1983).
RN [5]
RP PRELIMINARY SEQUENCE OF 267-397, AND SEQUENCE OF 1-6 FROM N.A.
RX MEDLINE=92196896; PubMed=6281737;
RA Dobson M.J., Tuite M.F., Roberts N.A., Kingsman A.J., Kingsman S.M.,
RA Perkins R.E., Conroy S.C., Dunbar B., Fothergill L.A.;
RT "Conservation of high efficiency promoter sequences in Saccharomyces
RT cerevisiae.";
RL Nucleic Acids Res. 10:2625-2637(1982).
RN [6]
RP SEQUENCE OF 149-160; 197-208 AND 244-255.
RX STRAIN-ATCC 44827 / SKQ2N;
RA MEDLINE=98170312; PubMed=9509572;
RA Norbeck J., Blomberg A.;
RT "Two-dimensional electrophoretic separation of yeast proteins using a
RT non-linear wide range (pH 3-10) immobilized pH gradient in the first
RT dimension: reproducibility and evidence for isoelectric focusing of
RT alkaline (pI > 7) proteins.";
RL Yeast 13:1519-1534(1997).
RN [7]
RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=94207923; PubMed=6765200;
RA Watson H.C., Walker N.P.C., Shaw P.J., Bryant T.N., Wendell P.L.,
RA Fothergill L.A., Perkins R.E., Conroy S.C., Dobson M.J., Tuite M.F.,
RA Kingsman A.J., Kingsman S.M.;
RT "Sequence and structure of yeast phosphoglycerate kinase.";
RL EMBO J. 1:1635-1640(1982).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF MUTANT GLN-65.
RX MEDLINE=96264793; PubMed=8672447;
RA McPhillips T.M., Hsu B.T., Sherman M.A., Mas M.T., Rees D.C.;
RT "Structure of the R65Q mutant of yeast 3-phosphoglycerate kinase
RT complexed with Mg-AMP-PNP and 3-phospho-D-glycerate.";
RL Biochemistry 35:4118-4127(1996).
RN [9]
RP MUTAGENESIS OF ARG-21.
RX MEDLINE=92331667; PubMed=1628654;
RA Walker P.A., Joao H.C., Littlechild J.A., Williams R.J.P.,
RA Watson H.C.;
RT "Characterisation of yeast phosphoglycerate kinase modified by
RT mutagenesis at residue 21.";
RL Eur. J. Biochem. 207:29-37(1992).
CC -1- CATALYTIC ACTIVITY: ATP + 3-phospho-D-glycerate = ADP + 3-
CC phospho-D-glyceroyl phosphate.

CC -1- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
CC -1- SUBUNIT: MONOMER.
CC -1- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
CC -1- CAUTION: THERE ARE CONFLICTS WITH THE SEQUENCE STORED IN PDB.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; J01342; AAA88729.1; -;
DR EMBL; X59720; CAA42329.2; -;
DR EMBL; M14438; AAA34864.1; -;
DR EMBL; K00553; AAA34863.1; ALT_SEQ.
DR PIR; A00570; KIBYG.
DR PIR; S05766; S05766.
DR PIR; S19422; S19422.
DR PDB; 3PGK; 15-APR-92.
DR PDB; 1QPG; 10-JUN-96.
DR COMPLETFAST-2DPAGE; P00560; -;
DR SGD; S0000605; PGK1.
DR InterPro; IPR001576; PGK.
DR Pfam; PF00162; PGK. 1.
DR PRINTS; PR00477; PHGLYCKINASE.
DR PROSITE; PS00111; PGLYCERATE_KINASE; 1.
KW Transferase; Kinase; Acetylation; Glycolysis; 3D-structure.
FT INIT_MET 0
FT MOD_RES 1 1 ACETYLATION.
FT CONFLICT 184 184 G -> S (IN REF. 3).
FT CONFLICT 190 190 E -> I (IN REF. 3).
FT STRAND 18 21
FT HELIX 37 52
FT STRAND 58 60
FT HELIX 77 87
FT STRAND 91 92
FT HELIX 100 108
FT STRAND 113 116
FT HELIX 142 152
FT TURN 153 153
FT STRAND 157 160
FT TURN 171 172
FT STRAND 181 183
FT HELIX 185 197
FT TURN 198 198
FT STRAND 205 209
FT TURN 216 216
FT STRAND 217 225
FT HELIX 231 234
FT STRAND 237 243
FT HELIX 246 251
FT TURN 257 258
FT HELIX 259 273
FT TURN 274 274
FT STRAND 276 277
FT STRAND 282 285
FT STRAND 295 298
FT HELIX 315 325
FT STRAND 330 333
FT STRAND 340 341
FT TURN 343 344
FT STRAND 348 359
FT TURN 360 361
FT STRAND 366 368
FT HELIX 372 380
FT TURN 383 384
FT STRAND 387 388
FT HELIX 393 401
FT TURN 402 402
FT HELIX 406 410
SQ SEQUENCE 415 AA; 44607 MW; 9E4E591546A82F40 CRC64;

```

Query Match          66.7%; Score 28; DB 1; Length 418;
Best Local Similarity 75.0%; Pred. No. 89;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```

RESULT 34			
ID	6PGD_LACLA	STANDARD;	PRT; 472 AA.
AC	Q9CHU6;		
DT	16-OCT-2001 (Rel. 40, Created)		
DT	16-OCT-2001 (Rel. 40, Last sequence update)		
DT	01-MAR-2002 (Rel. 41, Last annotation update)		
DE	6-phosphogluconate dehydrogenase (EC 1.1.1.44).		
GN	GND OR LL0622.		
OS	Lactococcus lactis (subsp. lactis) (Streptococcus lactis).		
OC	Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;		
OC	Lactococcus.		

OX NCBI_TaxID=1360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IL1403;
 RX MEDLINE=21235186; PubMed=11337471;
 RA Bolotin A., Winkler P., Mauger S., Jaillon O., Malarne K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium *Lactococcus*
 RT *lactis* ssp. *lactis* IL1403.";
 RL Genome Res. 11:731-753(2001).
 CC -1- CATALYTIC ACTIVITY: 6-phospho-D-gluconate + NADP(+) = D-ribulose
 CC 5-phosphate + CO(2) + NADPH.
 CC -1- PATHWAY: HEXOSE MONOPHOSPHATE SHUNT.
 CC -1- SIMILARITY: BELONGS TO THE 6-PHOSPHOGLUCONATE DEHYDROGENASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE006295; AAK04720.1; -;
 DR InterPro; IPR001744; 6PGD.
 DR Pfam; PF00393; 6PGD; 1.
 DR PRINTS; PR00076; 6PGDHRGNASE.
 DR PROSITE; PS00461; 6PGD; 1.
 KW Gluconate utilization; Oxidoreductase; Pentose shunt; NADP;
 KW Complete proteome.
 SQ SEQUENCE 472 AA; 52424 MW; 6EF4F504F217E55F CRC64;

 Query Match 66.7%; Score 28; DB 1; Length 472;
 Best Local Similarity 71.4%; Pred. No. 1e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

 QY 1 QRIKASK 7
 Db 297 ERVKASK 303
 :|:||||

 RESULT 35
 ID 6PGD_LACLC STANDARD; PRT; 472 AA.
 AC P96789;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 6-phosphogluconate dehydrogenase (EC 1.1.1.44).
 GN GND.
 OS *Lactococcus lactis* (subsp. *cremoris*) (*Streptococcus cremoris*).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC *Lactococcus*.
 ON NCBI_TaxID=1359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MG1363;
 RA Tetrad E., Hanau S., Wells J.M., Le Page R.W.F., Barrett M.P.;
 RT "6-phosphogluconate dehydrogenase from *Lactococcus lactis*: primary
 RT sequence and biochemical characterization.";
 RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
 CC -1- CATALYTIC ACTIVITY: 6-phospho-D-gluconate + NADP(+) = D-ribulose
 CC 5-phosphate + CO(2) + NADPH.
 CC -1- PATHWAY: HEXOSE MONOPHOSPHATE SHUNT.
 CC -1- SIMILARITY: BELONGS TO THE 6-PHOSPHOGLUCONATE DEHYDROGENASE
 CC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; U74322; AAC12804.1; -;
 DR HSP; P00349; 2PGD.
 DR InterPro; IPR001744; 6PGD.
 DR Pfam; PF00393; 6PGD; 1.
 DR PRINTS; PR00076; 6PGDHRGNASE.
 DR PROSITE; PS00461; 6PGD; 1.
 KW Gluconate utilization; Oxidoreductase; Pentose shunt; NADP.
 SQ SEQUENCE 472 AA; 52444 MW; 739958A068D63CD0 CRC64;

 Query Match 66.7%; Score 28; DB 1; Length 472;
 Best Local Similarity 71.4%; Pred. No. 1e+02;
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

 QY 1 QRIKASK 7
 Db 297 ERVKASK 303
 :|:||||

 RESULT 36
 ID YGZ9_YEAST STANDARD; PRT; 704 AA.
 AC P53061;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Hypothetical 82.8 kDa protein in HPML-PDB1 intergenic region.
 GN YGL249W OR NRC704.
 OS *Saccharomyces cerevisiae* (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 ON NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=S288C / FY1679;
 RX MEDLINE=97127827; PubMed=8972578;
 RA Coissac E., Maillier E., Robineau S., Netter P.;
 RT "Sequence of a 39,411 bp DNA fragment covering the left end of
 RT chromosome VII of *Saccharomyces cerevisiae*.";
 RL Yeast 12:1555-1562(1996).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X94357; CAA64138.1; -;
 DR EMBL; Z72771; CAA96969.1; -;
 DR SGD; S0003218; ZIP2.
 KW Hypothetical protein.
 SQ SEQUENCE 704 AA; 82816 MW; 171387494869B56E CRC64;

 Query Match 66.7%; Score 28; DB 1; Length 704;
 Best Local Similarity 66.7%; Pred. No. 1.5e+02;
 Matches 6; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

 QY 1 QRIKASKSF 9
 Db 222 QRISPEKSF 230
 :|:||||

 RESULT 37
 ID TOPI_MYCPN STANDARD; PRT; 711 AA.
 AC P78032;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE DNA topoisomerase I (EC 5.99.1.2) (Omega-protein) (Relaxing enzyme)
DE (Unwinding enzyme) (Swivelase).
GN TOPA OR MPN261 OR MP572.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfreich R., Hilbert H., Plagens H., Pirkil E., Li B.-C.,
RA Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae";
RL Nucleic Acids Res. 24:4420-4449(1996).
CC -!- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -!- CATALYTIC ACTIVITY: ATP-independent breakage of single-stranded
CC DNA, followed by passage and rejoining.
CC -!- SUBUNIT: MONOMER (BY SIMILARITY).
CC -!- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -!- SIMILARITY: BELONGS TO PROKARYOTIC TYPE I/III TOPOISOMERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000056; AAB96220.1; -
DR HSSP; P06612; 1ECL.
DR InterPro; IPRO03601; DNATopI_ATP_bind.
DR InterPro; IPRO03602; DNATopI_DNA_bind.
DR InterPro; IPRO00380; Pro_topoisomerase.
DR InterPro; IPRO02936; Toprim.
DR Pfam; PF01131; Topoisom_bac; 1.
DR Pfam; PF01751; Toprim; 1.
DR Pfam; PF01396; zf-C4_Topoiso; 2.
DR PRINTS; PR00417; PRTPISMRASE1.
DR SMART; SM00437; TOP1AC; 1.
DR SMART; SM00436; TOP1BC; 1.
DR SMART; SM00493; TOPRIM; 1.
DR PROSITE; PS00396; TOPOISOMERASE_I_PROK; 1.
KW Isoomerase; Topoisomerase; DNA-binding; Zinc-finger; Metal-binding;
KW Repeat; Complete proteome.
FT ZN_FING 624 652 C4-TYPE 1.
FT ZN_FING 673 702 C4-TYPE 2.
FT ACT_SITE 340 340 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 711 AA; 81965 MW; 16B863588B06582E CRC64;

Query Match 66.7%; Score 28; DB 1; Length 711;
Best Local Similarity 55.8%; Pred. No. 1.5e+02;
Matches 5; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
||:| | |
Db 611 QRVKKEKF 619.

RESULT 38
ATX7_HUMAN
ID ATX7_HUMAN STANDARD; PRT; 892 AA.
AC O15265; O75329; O75328; O9Y6P8;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Ataxin 7 (Spinocerebellar ataxia type 7 protein).
GN SCA7.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lymphoblast;
RX MEDLINE=97434213; PubMed=9288099;
RA David G., Abbas N., Stevanin G., Duerr A., Yvert G., Cancel G.,
RA Weber C., Imbert G., Saudou F., Antoniou E., Drabkin H., Gemmill R.,
RA Giunti P., Benomar A., Wood N., Ruberg M., Agid Y., Mandel J.-L.,
RA Brice A.;
RT "Cloning of the SCA7 gene reveals a highly unstable CAG repeat
RT expansion.";
RL Nat. Genet. 17:65-70(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RX MEDLINE=98087568; PubMed=9425224;
RA Del-Favero J., Krols L., Michalik A., Theuns J., Loeffgren A.,
RA Goossens D., Wehnert A., Van den Bossche D., Van Zand K.,
RA Backhovens H., van Regenmortel N., Martin J.-J., Van Broeckhoven C.;
RT "Molecular genetic analysis of autosomal dominant cerebellar ataxia
RT with retinal degeneration (ADCA type II) caused by CAG triplet repeat
RT expansion.";
RL Hum. Mol. Genet. 7:177-186(1998).
RN [3]
RP SUBCELLULAR LOCATION.
RX MEDLINE=99371772; PubMed=10441328;
RA Kaytor M.D., Duvick L.A., Skinner P.J., Koob M.D., Ranum L.P.,
RA Orr H.T.;
RT "Nuclear localization of the spinocerebellar ataxia type 7 protein,
RT ataxin-7";
RL Hum. Mol. Genet. 8:1657-1664(1999).
CC -!- SUBCELLULAR LOCATION: NUCLEAR. IN ADDITION TO A DIFFUSE
CC DISTRIBUTION THROUGHOUT THE NUCLEUS, IT IS ASSOCIATED WITH THE
CC NUCLEAR MATRIX AND THE NUCLEOLUS.
CC -!- POLYMORPHISM: THE POLY-GLN REGION OF SCA7 IS HIGHLY POLYMORPHIC (4
CC 38-130 REPEATS) IN THE NORMAL POPULATION AND IS EXPANDED TO ABOUT
CC 35 REPEATS ARE PRONE TO FURTHER EXPANSION.
CC -!- DISEASE: DEFECTS IN SCA7 ARE THE CAUSE OF SPINOCEREBELLAR ATAXIA
CC TYPE 7; ALSO KNOWN AS AUTOSOMAL DOMINANT CEREBELLAR ATAXIA TYPE II
CC (ADCA-II). ADCA-II IS CHARACTERIZED BY PIGMENTARY MACULAR
CC DYSTROPHY AND RETINAL DEGENERATION LEADING TO BLINDNESS.
CC -!- SIMILARITY: BELONGS TO THE ATAXIN 7 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AJ000517; CA04154.1; -
DR EMBL; AF032105; AAC39765.1; -
DR EMBL; AF032102; AAC19162.1; -
DR EMBL; AF032103; AAC19163.1; -
DR MIM; 164500; -
DR InterPro; IPRO00822; znf-C2H2.
DR SMART; SM00355; ZNF_C2H2; 1.
KW Polymorphism; Triplet repeat expansion; Disease mutation;
KW Nuclear protein.
FT DOMAIN 16 20 POLY-ALA.
FT DOMAIN 23 28 POLY-ALA.
FT DOMAIN 30 39 POLY-GLN.
FT DOMAIN 40 45 POLY-PRO.
FT DOMAIN 51 55 POLY-PRO.
FT DOMAIN 171 174 POLY-SER.
FT DOMAIN 213 219 POLY-SER.

```

FT DOMAIN 647 654 POLY-SER.
FT DOMAIN 717 730 POLY-SER.
FT DOMAIN 840 845 POLY-SER.
FT VARIANT 264 264 K -> R (IN DBSNP:10533338).
FT VARIANT 663 663 /FTID=VAR_011823.
FT VARIANT 105 105 /FTID=VAR_011824.
FT CONFLICT 129 129 P -> H (IN REF. 2; AAC19162).
FT CONFLICT 129 129 C -> S (IN REF. 2; AAC19162).
FT CONFLICT 862 862 V -> M (IN REF. 2).
FT CONFLICT 888 892 PKARP -> VGNGL (IN REF. 2; AAC39765 AND AF032103).
FT SQ SEQUENCE 892 AA; 95451 MW; 9AEA787D77103C5F CRC64;

Query Match 66.7%; Score 28; DB 1; Length 892;
Best Local Similarity 62.5%; Pred. No. 2e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8
Db 672 QKLKSKS 679

RESULT 39
LAFX_LACJO STANDARD; PRT; 62 AA.
AC Q48509;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Bacteriocin lactacin F, subunit lafx precursor.
GN LAFX.
OS Lactobacillus johnsonii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=33959;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=ATCC 11506 / VPI 11088;
RX MEDLINE=94113718; PubMed=8285694;
RA Fremaux C., Ahn C., Klaenhammer T.R.;
RT "Molecular analysis of the lactacin F operon.";
RA Appl. Environ. Microbiol. 59:3906-3915(1993).
RL -1- FUNCTION: HEAT STABLE BACTERIOCIN ACTIVE AGAINST ENTEROCOCCUS
CC FAECALIS AND OTHER LACTOBACILLI.
CC -1- SUBUNIT: THIS BACTERIOCIN DEPENDS UPON THE COMPLEMENTATION OF TWO
CC PEPTIDES FOR ACTIVITY: LAFB AND LAFX. ASSOCIATED WITH A 180 Kda
CC BACTERIOCIN COMPLEX.
CC -1- SIMILARITY: BELONGS TO THE CLASS IIB BACTERIOCINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M57961; AAA16637.1; -
CC Antibiotic; Bacteriocin.
KW PROPEP 1 14 BY SIMILARITY.
FT CHAIN 15 62 BACTERIOCIN LACTACIN F, SUBUNIT LAFX.
FT SEQUENCE 62 AA; 6250 MW; B8EB8BE795830716 CRC64;

Query Match 64.3%; Score 27; DB 1; Length 62;
Best Local Similarity 85.7%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 3 IKASKSF 9
Db 32 IKACKSF 38

```

```

RESULT 40
RS17_BACHD STANDARD; PRT; 86 AA.
AC Q929X5; Q9JXP8.
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S17.
GN RPSQ OR BH0143.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RX MEDLINE=99209008; PubMed=10192928;
RA Takami H., Takaki Y., Nakasone K., Hirama C., Inoue A., Horikoshi K.;
RT "Sequence analysis of a 32-kb region including the major ribosomal
RT protein gene clusters from alkaliphilic Bacillus sp. strain C-125.";
RL Biosci. Biotechnol. Biochem. 63:452-455(1999).
RN [2]
RC SEQUENCE FROM N.A.
RX STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
RC -1- FUNCTION: PROTEIN S17 BINDS SPECIFICALLY TO THE 5' END OF 16S
CC RIBOSOMAL RNA (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE S17P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AB017508; BAA75280.1; -
CC HSP; P23828; IRIP.
CC InterPro: IPR000266; Ribosomal_S17.
CC Pfam: PF00366; Ribosomal_S17; 1.
CC PRINTS; PR00973; RIBOSOMALS17.
CC ProDom: PD001295; Ribosomal_S17; 1.
CC PROSITE; PS00056; RIBOSOMAL_S17; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
SQ SEQUENCE 86 AA; 10182 MW; 16004E0F6B643954 CRC64;

Query Match 64.3%; Score 27; DB 1; Length 86;
Best Local Similarity 55.6%; Pred. No. 28;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
Db 37 KRVYKSKF 45

RESULT 41
RS17_BACSU STANDARD; PRT; 86 AA.
AC P12874;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE 30S ribosomal protein S17 (BS16).
GN RPSQ.
OS Bacillus subtilis.

```

```
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=90016806; PubMed=2508062;
RA Henkin T.M., Moon S.H., Mattheakis L.C., Nomura M.;
RT "Cloning and analysis of the spc ribosomal protein operon of Bacillus
RT subtilis: comparison with the spc operon of Escherichia coli.";
RL Nucleic Acids Res. 17:7469-7486(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / MARBURG;
RX MEDLINE=96186897; PubMed=8635744;
RA Suh J.W., Boylan S.A., Oh S.H., Price C.W.;
RT "Genetic and transcriptional organization of the Bacillus subtilis
RT spc-alpha region.";
RL Gene 169:17-23(1996).
RN [3]
RP SEQUENCE OF 1-37.
RX MEDLINE=82219212; PubMed=6806564;
RA Higo K.I., Otake E., Osawa S.;
RT "Purification and characterization of 30S ribosomal proteins from
RT Bacillus subtilis: correlation to Escherichia coli 30S proteins.";
RL Mol. Genet. 185:239-244(1982).
CC -!- FUNCTION: PROTEIN S17 BINDS SPECIFICALLY TO THE 5' END OF 16S
CC RIBOSOMAL RNA.
CC -!- SIMILARITY: BELONGS TO THE S17P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X15664; CA337700.1; -.
DR EMBL; L47971; AAB06808.1; -.
DR EMBL; Z99104; CAB11901.1; -.
DR PIR; S05991; R3BS17.
DR PIR; S11367; S11367.
DR HSSP; P23828; IRIIP.
DR Subtilisin; BG10757; rpsQ.
DR InterPro; IPR002666; Ribosomal_S17.
DR Pfam; PF00366; Ribosomal_S17; 1.
DR PRINTS; PR00973; RIBOSOMALS17.
DR PRODOM; PD001295; Ribosomal_S17; 1.
DR PROSITE; PS00056; RIBOSOMAL_S17; 1.
KW Ribosomal protein; rRNA-binding; Complete proteome.
FT INT_MET 0 2
FT CONFLICT 2 2 E -> ECE (IN REF. 2).
FT CONFLICT 23 23 T -> G (IN REF. 3).
SQ SEQUENCE 86 AA; 10068 MW; 9CADAA4D0F24BFA3 CRC64;

Query Match 64.3%; Score 27; DB 1; Length 86;
Best Local Similarity 55.6%; Pred. No. 28;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
Db 37 KRVYSKKF 45

RESULT 42
TVC2_MOUSE
ID TVC2_MOUSE STANDARD; PRT; 115 AA.
AC P03978;
DT 23-OCT-1986 (Rel. 02, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE T-cell receptor gamma chain V region V108B precursor (Fragment).
```

```
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85099339; PubMed=3917858;
RA Hayday A.C., Saito H., Gillies S.D., Kranz D.M., Tanigawa G.,
RA Eisen H.N., Tonesgawa S.;
RT "Structure, organization, and somatic rearrangement of T cell gamma
RT genes.";
RL Cell 40:259-269(1985).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M12832; AAA40307.1; -.
DR PIR; A02020; RWSVB.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; IgV; 1.
KW T-cell; Receptor; Signal.
FT SIGNAL 1 18
FT CHAIN 19 >115 T-CELL RECEPTOR GAMMA CHAIN V REGION
FT V108B.
FT DOMAIN 19 115
FT NON_TER 115 115 V SEGMENT.
FT SEQUENCE 115 AA; 13495 MW; 225D6095D96945A0 CRC64;

Query Match 64.3%; Score 27; DB 1; Length 115;
Best Local Similarity 55.6%; Pred. No. 39;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
Db 81 KKIEASKDF 89

RESULT 43
RNPA_LACLA
ID RNPA_LACLA STANDARD; PRT; 117 AA.
AC Q9CJ73;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Ribonuclease P protein component (EC 3.1.26.5) (RNaseP protein)
DE (RNase P protein) (Protein C5).
GN RNPA OR L10128.
OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Lactococcus.
OX NCBI_TaxID=1360;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IL1403;
RX MEDLINE=21235186; PubMed=11337471;
RA Bolotin A., Wincker P., Manger S., Jaillon O., Malarre K.,
RA Weissenbach J., Ehrlich S.D., Sorokin A.;
RT "The complete genome sequence of the lactic acid bacterium Lactococcus
RT lactis ssp. lactis IL1403.";
RL Genome Res. 11:731-753(2001).
CC -!- FUNCTION: RNaseP catalyzes the removal of the 5'-leader sequence
CC from pre-tRNA to produce the mature 5' terminus. It can also
CC cleave other RNA substrates such as 4.5S RNA. The protein
CC component plays an auxiliary but essential role in vivo by binding
CC to the 5'-leader sequence and broadening the substrate specificity
CC of the ribozyme (By similarity).
```

CC -1- CATALYTIC ACTIVITY: Endonucleolytic cleavage of RNA, removing 5'-
CC extra-nucleotide from tRNA precursor.
CC -1- SUBUNIT: Consists of a catalytic RNA component (M1 or rnpB) and a
CC protein subunit (By similarity).
CC -1- SIMILARITY: BELONGS TO THE RNPA FAMILY.

CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL: AE006251; AAK04226.1; -;
DR InterPro: IPR000100; Ribonuclease_P.
DR Pfam: PF00825; Ribonuclease_P; 1.
DR PROSITE: PS00648; RIBONUCLEASE_P; 1.
KW Hydrolyase; Nuclease; Endonuclease; tRNA processing; RNA-binding;
KW Complete proteome.
SQ SEQUENCE 117 AA; 13648 MW; 4F1001A7B4C63C96 CRC64;

Query Match 64.3%; Score 27; DB 1; Length 117;
Best Local Similarity 62.5%; Pred. No. 39;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 RIKASKSF 9
|:|:|:|:|
Db 8 RVRKSKDF 15

RESULT 44

TVC3_MOUSE
ID TVC3_MOUSE STANDARD; PRT; 132 AA.
AC P06324;
DT 01-JAN-1988 (Rel. 06, Created)
DT 01-JAN-1988 (Rel. 06, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE T-cell receptor gamma chain V region DFL12 precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85137888; PubMed=3871915;
RA Kranz D.M., Saito H., Heller M., Takagaki Y., Haas W., Eisen H.N.,
RA Toneygawa S.;
RT "Limited diversity of the rearranged T-cell gamma gene.";
RL Nature 313:752-755(1985).
DR PIR: A02019; RWS12.
DR HSP; P01709; 2MCG.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
KW T-cell; Receptor; Signal.
FT SIGNAL 1 18
FT CHAIN 19 132 T-CELL RECEPTOR GAMMA CHAIN V REGION
FT DFL12.
FT DOMAIN 19 114 V SEGMENT.
FT DOMAIN 115 132 J SEGMENT.
FT NON_TER 132 132
SQ SEQUENCE 132 AA; 15434 MW; 855D714D0A43CCFE CRC64;

Query Match 64.3%; Score 27; DB 1; Length 132;
Best Local Similarity 55.6%; Pred. No. 45;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
::|:|:|:|
Db 82 KKIEASKDF 90

RESULT 45
TVC1_MOUSE
ID TVC1_MOUSE STANDARD; PRT; 135 AA.
AC P01740;
DT 21-JUL-1986 (Rel. 01, Created)
DT 23-OCT-1986 (Rel. 02, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE T-cell receptor gamma chain V region V108A precursor.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85099339; PubMed=3917858;
RA Hayday A.C., Saito H., Gillies S.D., Kranz D.M., Tanigawa G.,
RA Eisen H.N., Toneygawa S.;
RT "Structure, organization, and somatic rearrangement of T cell gamma
RT genes.";
RL Cell 40:259-269(1985).
RN [2]
RP SEQUENCE FROM N.A. (CLONE PHDS4/PHDS203).
RX STRAIN=BALB.B;
RX MEDLINE=84245824; PubMed=6330561;
RA Saito H., Kranz D.M., Takagaki Y., Hayday A.C., Eisen H.N.,
RA Toneygawa S.;
RT "Complete primary structure of a heterodimeric T-cell receptor
RT deduced from cDNA sequences.";
RL Nature 309:757-762(1984).
CC -1- MISCELLANEOUS: THIS MRNA WAS ISOLATED FROM A CYTOTOXIC
CC T-LYMPHOCYTE. THE GENE CORRESPONDING TO THIS MRNA IS REARRANGED
CC SPECIFICALLY IN T CELLS AND ITS ORGANIZATION IS SIMILAR TO AN IG
CC GENE, WITH V (OR V+D), J, AND C REGIONS.
DR PIR: A02018; RWSV1.
DR HSP; P01703; 7FAB.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR003596; Ig_v.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00406; IGV; 1.
KW T-cell; Receptor; Signal.
FT SIGNAL 1 18
FT CHAIN 19 135 T-CELL RECEPTOR GAMMA CHAIN V REGION
FT V108A.
FT DOMAIN 19 116 V SEGMENT.
FT DOMAIN 117 135 J SEGMENT.
FT NON_TER 135 135
SQ SEQUENCE 135 AA; 15930 MW; 402145C95E924D7D CRC64;

Query Match 64.3%; Score 27; DB 1; Length 135;
Best Local Similarity 55.6%; Pred. No. 46;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
::|:|:|:|
Db 82 KKIEASKDF 90

Search completed: November 6, 2002, 12:09:22
Job time : 8.33333 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:16 ; Search time 11 Seconds
(without alignments)
78.619 Million cell updates/sec

Title: US-09-905-083-99
Perfect score: 42
Sequence: 1 QRIKASKSF 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_71:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	253	2 A53968	serine proteinase
2	34	81.0	373	2 D89004	protein T24A6.11
3	33	78.6	427	2 A49518	kallistatin precursor
4	32	76.2	895	2 D72571	probable DNA-dir
5	31	73.8	248	2 A40625	tetrachloro-p-hydr
6	31	73.8	298	2 S32368	beta-SNAP protein
7	31	73.8	302	2 T50147	mak16-like protein
8	31	73.8	362	2 F81409	probable phosphatase
9	30	71.4	142	2 C39610	BET1 protein - yea
10	30	71.4	321	2 C71463	probable p-loop AR
11	30	71.4	331	2 H97051	probable dehydroge
12	30	71.4	427	2 T48159	hypothetical prote
13	30	71.4	517	2 T40129	hypothetical prote
14	30	71.4	627	2 C86850	neutral endopeptid
15	30	71.4	627	2 A47098	lactococcal endope
16	30	71.4	627	2 F53290	endopeptidase Pepo
17	30	71.4	875	2 T19678	hypothetical prote
18	30	71.4	985	2 D82776	pyruvate dehydroge
19	30	71.4	1102	2 H84545	probable ubiquitin
20	30	71.4	3660	1 S02041	dystrophin, muscle
21	29	69.0	65	2 AF2269	hypothetical prote
22	29	69.0	141	1 B46322	16K protein - toba
23	29	69.0	172	2 H81251	NADH dehydrogenase
24	29	69.0	182	2 C64176	hypothetical prote
25	29	69.0	203	2 C90678	thiogalactoside ac
26	29	69.0	203	2 G85528	thiogalactoside ac
27	29	69.0	268	2 E84339	quinolinate phosph
28	29	69.0	349	2 AE2568	WD repeat protein
29	29	69.0	368	2 T22587	hypothetical prote

30	29	69.0	463	2 G97300	aspartyl/asparagin
31	29	69.0	484	2 JC5282	phosphogluconate d
32	29	69.0	526	1 G71081	probable helicase
33	29	69.0	563	2 B82883	hypothetical prote
34	29	69.0	584	2 T48273	hypothetical prote
35	29	69.0	585	2 I37216	calicin - human (f
36	29	69.0	669	2 AB2028	hypothetical prote
37	29	69.0	691	2 T46476	hypothetical prote
38	29	69.0	717	2 AC1419	DNA topoisomerase
39	29	69.0	738	2 E86294	hypothetical prote
40	29	69.0	743	2 C86168	hypothetical prote
41	29	69.0	759	2 S67164	probable membrane
42	29	69.0	940	2 A40985	projectin - fruit
43	29	69.0	1018	2 T15297	hypothetical prote
44	29	69.0	1742	2 S24600	projectin - fruit
45	29	69.0	6658	2 T13931	projectin - fruit

ALIGNMENTS

RESULT 1

A53968 serine proteinase SCCE precursor - human
N:Alternate names: stratum corneum chymotryptic enzyme
C:Species: Homo sapiens (man)
C:Date: 07-Jul-1995 #sequence_revision 07-Jul-1995 #text_change 22-Jun-1999
C:Accession: A53968
R: Hansson, L.; Stroemqvist, M.; Baeckman, A.; Wallbrandt, P.; Carlstein, A.; Egelrud, J. Biol. Chem. 269, 19420-19426, 1994
A: Title: Cloning, expression, and characterization of stratum corneum chymotryptic er
A: Reference number: A53968; MUID: 94308225
A: Accession: A53968
A: Status: preliminary
A: Molecule type: mRNA
A: Residues: 1-253 <HAN>
A: Cross-references: GB:L33404; NID: g521214; PIDN: AAC37551.1; PID: g532504
C: Genetics:
A: Gene: GDB: PRSS6; SCCE
A: Cross-references: GDB: 377730
A: Map position: 7q35-7q35
C: Superfamily: trypsin; trypsin homology
F: 30-245/Domain: trypsin homology <TRY>

Query Match 100.0%; Score 42; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.085;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 91 QRIKASKSF 99

RESULT 2

D89004 protein T24A6.11 [imported] - Caenorhabditis elegans
C: Species: Caenorhabditis elegans
C: Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 10-May-2001
C: Accession: D89004
R: Anonymous, The C. elegans Sequencing Consortium.
Science 282, 2012-2018, 1998
A: Title: Genome sequence of the nematode C. elegans: a platform for investigating bir
A: Reference number: A75000; MUID: 99069613; PMID: 9851916
A: Note: see websites genome.wustl.edu/gsc/c_elegans/ and www.sanger.ac.uk/Projects/C
A: Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999;
A: Accession: D89004
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-373 <STO>
A: Cross-references: GB:chr_V; PIDN: AAC17798.1; PID: g3168946; GSPDB: GN00023; CESP: T24.
A: Note: contains similarity to C4-type zinc fingers
C: Genetics:
A: Gene: T24A6.11

A:Map position: 5

Query Match 81.0%; Score 34; DB 2; Length 373;
Best Local Similarity 77.8%; Pred. No. 8.6;
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9

Db 9 OEIKSKSF 17

RESULT 3

A49516

kallistatin precursor - human

N:Alternate names: proteinase inhibitor 4; tissue kallikrein inhibitor

C:Species: Homo sapiens (man)

C:Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 26-May-2000

C:Accession: A49518; A55560; A45093

R:Chai, K.X.; Chen, L.M.; Chao, J.; Chao, L.

J. Biol. Chem. 268, 24498-24505, 1993

A:Title: Kallistatin: a novel human serine proteinase inhibitor. Molecular cloning, tissue

A:Reference number: A49518; MUID:94043294

A:Accession: A49518

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-427 <CHA>

A:Cross-references: GB:L19684; NID:g425145; PIDN:AAA59454.1; PID:g425146

R:Chai, K.X.; Ward, D.C.; Chao, J.; Chao, L.

Genomics 23, 370-378, 1994

A:Title: Molecular cloning, sequence analysis, and chromosomal localization of the human

A:Reference number: A55560; MUID:95137583

A:Accession: A55560

A>Status: preliminary; not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-427 <CH2>

A:Cross-references: GB:L28101; NID:g609489; PIDN:AAC41706.1; PID:g619793

R:Zhou, G.X.; Chao, L.; Chao, J.

J. Biol. Chem. 267, 25873-25880, 1992

A:Title: Kallistatin: a novel human tissue kallikrein inhibitor. Purification, character

A:Reference number: A45093; MUID:93100304

A:Accession: A45093

A:Molecule type: protein

A:Residues: 389-403 <ZHO>

A:Cross-references: PIDN:AB24557.1; PID:g261993

A:Note: sequence extracted from NCBI backbone (NCBIP:120847)

C:Genetics:

A:Gene: GDB:PI4; KLST

A:Cross-references: GDB:266537; OMIM:147935

A:Map position: 14q31-14q32.1

C:Superfamily: antithrombin III

C:Keywords: blocked amino end; glycoprotein; pyroglutamic acid; serine proteinase inhibi

F:1-20/Domain: signal sequence #status predicted <SIG>

F:21-427/Product: kallistatin #status predicted <MAT>

F:21/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted

F:33,109,157,238/Binding site: carboxylate (Asn) (covalent) #status predicted

F:388/Inhibitory site: Phe (tissue kallikrein) #status experimental

Query Match 78.6%; Score 33; DB 2; Length 427;

Best Local Similarity 66.7%; Pred. No. 17;

Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9

Db 355 QKLEASKSF 17

RESULT 4

D72571

probable DNA-directed RNA polymerase subunit A' APE1853 - Aeropyrum pernix (strain K1)

C:Species: Aeropyrum pernix

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: D72571

R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Haikawa, Y.; Jin-no, K.; Takah

awa, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.
DNA Res. 6, 83-101, 1999

A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero

A:Reference number: A72450; MUID:99310339

A:Accession: D72571

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-895 <KAW>

A:Cross-references: DDBJ:AP000062; NID:g5105244; PIDN:BAA80857.1; PID:g5105544

A:Experimental source: strain K1

C:Genetics:

A:Gene: APE1853

C:Superfamily: Halobacterium DNA-directed RNA polymerase chain A

Query Match 76.2%; Score 32; DB 2; Length 895;

Best Local Similarity 75.0%; Pred. No. 60;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8

Db 885 QRVKASKA 892

RESULT 5

A40625

tetrachloro-p-hydroquinone reductive dehalogenase (EC 1.-.-.-) - Flavobacterium sp.

N:Alternate names: glutathione S-transferase homolog; TeCH reductive dehalogenase

C:Species: Flavobacterium sp.

C:Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 08-Oct-1999

C:Accession: A40625

R:Orser, C.S.; Dutton, J.; Lange, C.; Jablonski, P.; Xun, L.; Hargis, M.

J. Bacteriol. 175, 2640-2644, 1993

A:Title: Characterization of a Flavobacterium glutathione S-transferase gene involved

A:Reference number: A40625; MUID:93239690

A:Accession: A40625

A>Status: preliminary

A:Molecule type: nucleic acid

A:Residues: 1-248 <ORS>

A:Cross-references: GB:M98559; NID:g148689; PIDN:AAA24921.1; PID:g148690

A:Experimental source: ATCC 39723

A:Note: sequence extracted from NCBI backbone (NCBIN:130432, NCBIP:130433)

C:Keywords: oxidoreductase

Query Match 73.8%; Score 31; DB 2; Length 248;

Best Local Similarity 66.7%; Pred. No. 28;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9

Db 227 QRMKARRSF 235

RESULT 6

S32368

beta-SNAP protein - bovine

C:Species: Bos primigenius taurus (cattle)

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 10-Dec-1999

C:Accession: S32368

R:Whiteheart, S.W.; Griff, I.C.; Brunner, M.; Clary, D.O.; Mayer, T.; Buhrow, S.A.; R

Nature 362, 353-355, 1993

A:Title: SNAP family of NSF attachment proteins includes a brain-specific isoform.

A:Reference number: S32367; MUID:93205122

A:Accession: S32368

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-298 <WHI>

C:Superfamily: alpha-soluble NSF attachment protein

Query Match 73.8%; Score 31; DB 2; Length 298;

Best Local Similarity 66.7%; Pred. No. 33;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9

Db 19 KRVKASHSF 27
:|:|||||

RESULT 7

T50147
mak16-like protein [imported] - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 21-Jul-2000
C:Accession: T50147
R:Hamlin, N.; Churcher, C.M.; McDougall, R.C.; Rajandream, M.A.; Barrell, B.G.
submitted to the EMBL Data Library, November 1999
A:Reference number: Z25043
A:Accession: T50147
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-302 <HAM>
A:Cross-references: EMBL:AL132798; PIDN:CAB60698.1; GSPDB:GN00066; SPDB:SPAC222.06
A:Experimental source: strain 972h(-); cosmid c222
C:Genetics:
A:Gene: SPDB:SPAC222.06
A:Map position: 1
A:Introns: 21/2
C:Superfamily: MAK16 protein

Query Match 73.8%; Score 31; DB 2; Length 302;

Best Local Similarity 66.7%; Pred. No. 34;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9

|||||

Db 78 ORIKLSKNY 86

RESULT 8

F81409
probable phosphate transport system permease protein Cj0615 [imported] - Campylobacter jejuni
C:Species: Campylobacter jejuni
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C:Accession: F81409
R:Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chillingworth, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barrett, N.; et al. 2000
Nature 403, 665-668, 2000
A:Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervirulence
A:Reference number: A81250; MUID:20150912
A:Accession: F81409
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-362 <PAR>
A:Cross-references: GB:AL139075; GB:AL111168; NID:g6967817; PIDN:CAB75251.1; PID:g696808
A:Experimental source: serotype O2, strain NCTC 11168
C:Genetics:
A:Gene: pstA; Cj0615

Query Match 73.8%; Score 31; DB 2; Length 362;

Best Local Similarity 77.8%; Pred. No. 40;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9

:|:|||||

Db 7 KROKASKSF 15

RESULT 9

C39610
BET1 protein - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YIA4c; protein YIL004c; SLY12 protein
C:Species: Saccharomyces cerevisiae
C:Date: 13-Sep-1991 #sequence_revision 13-Sep-1991 #text_change 29-Oct-1999
C:Accession: C39610; S48453; S58677
R:Bascher, C.; Ossig, R.; Gallwitz, D.; Schmitt, H.D.
Mol. Cell. Biol. 11, 872-885, 1991
A:Title: Identification and structure of four yeast genes (SLY) that are able to suppress

A:Reference number: A39610; MUID:91117254
A:Accession: C39610
A:Molecule type: DNA
A:Residues: 1-142 <DAS>
A:Cross-references: EMBL:X54237; NID:g4483; PIDN:CAA38143.1; PID:g4484
R:Rowley, N.

submitted to the EMBL Data Library, August 1994

A:Reference number: S48442

A:Accession: S48453

A:Molecule type: DNA

A:Residues: 1-142 <ROW>

A:Cross-references: GB:Z47047; EMBL:Z38113; NID:g603997; PID:g763342; MIPS:YIL004c
R:Voss, H.; Tamames, J.; Teodoru, C.; Valencia, A.; Sensen, C.; Wiemann, S.; Schwaiger
Yeast 11, 61-78, 1995

A:Title: Nucleotide sequence and analysis of the centromeric region of yeast chromosome

A:Reference number: S50795; MUID:95282515

A:Accession: S58677

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-142 <VOS>

A:Cross-references: EMBL:X79743

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1994

C:Genetics:

A:Gene: SGD:BET1; SLY12

A:Cross-references: SGD:S0001266; MIPS:YIL004c

A:Map position: 9L

A:Introns: 4/2

C:Keywords: transmembrane protein

F:126-142/Domain: transmembrane #status predicted <TMM>

Query Match 71.4%; Score 30; DB 2; Length 142;

Best Local Similarity 87.5%; Pred. No. 26;

Matches 7; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKS 8

|||||

Db 67 QRIKALKS 74

RESULT 10

C71463
probable p-loop ATPase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: C71463
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, D.; et al. 1998
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
A:Reference number: A71570; MUID:99000809
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-321 <ARN>
A:Cross-references: GB:AE001357; GB:AE001273; NID:g3329311; PIDN:AAC68437.1; PID:g3329311
A:Experimental source: serotype D, strain UW-3/Cx
C:Genetics:
A:Gene: mesJ

Query Match 71.4%; Score 30; DB 2; Length 321;

Best Local Similarity 66.7%; Pred. No. 61;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9

:|:|||||

Db 277 ERLVASKSF 285

RESULT 11

H97051
probable dehydrogenase, YULF B. subtilis ortholog [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001
C:Accession: H97051

R:Nolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum ATCC824
A:Reference number: A96900; MUID:21359325; PMID:21359325
A:Accession: H97051
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-331 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK79203.1; PID:gl5024156; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics: CAC1231
C:Superfamily: Escherichia coli yjyR protein

Query Match 71.4%; Score 30; DB 2; Length 331;
Best Local Similarity 75.0%; Pred. No. 63;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8
||| |||
Db 303 QRLKVS 310

RESULT 12
T48159
hypothetical protein T1008.80 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48159
R:Bevan, M.; Pohl, T.; Welzenegger, T.; Bancroft, I.; Mewes, H.W.; Lemcke, K.; Mayer, K.
submitted to the Protein Sequence Database, March 2000
A:Reference number: 224486
A:Accession: T48159
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-427 <BEV>
A:Cross-references: EMBL:AL161746
A:Experimental source: cultivar Columbia; BAC clone T1008
C:Genetics:
A:Map position: 5
A:Introns: 317/1
A:Note: T1008.80

Query Match 71.4%; Score 30; DB 2; Length 427;
Best Local Similarity 66.7%; Pred. No. 81;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
: ||| |||
Db 140 RVRKASVSF 148

RESULT 13
T40129
hypothetical protein SPBC2F12.12c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T40129
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Skelton, J.; Churcher, C.M.
submitted to the EMBL data Library, June 1997
A:Reference number: 221907
A:Accession: T40129
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-517 <WOO>
A:Cross-references: EMBL:297211; PIDN:CAB10159.1; GSPDB:GN00067; SPDB:SPBC2F12.12c
A:Experimental source: strain 972h; cosmid c2F12
C:Genetics:
A:Gene: SPDB:SPBC2F12.12c
A:Map position: 2
A:Introns: 42/3; 126/3; 157/3; 308/3

Query Match 71.4%; Score 30; DB 2; Length 517;
Best Local Similarity 77.8%; Pred. No. 98;
Matches 7; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
||| |||
Db 187 QRIAPKSF 195

RESULT 14
C86850
neutral endopeptidase [imported] - Lactococcus lactis subsp. lactis (strain IL1403)
C:Species: Lactococcus lactis subsp. lactis
C:Date: 23-Mar-2001 #sequence_revision 23-Mar-2001 #text_change 03-Aug-2001
C:Accession: C86850
R:Polotin, A.; Wincker, P.; Mauger, S.; Jaillon, O.; Malarme, K.; Weissenbach, J.; Eh
Genome Res. 11, 731-753, 2001
A:Title: The complete genome sequence of the lactic acid bacterium Lactococcus lactis
A:Reference number: A86625; MUID:21235186; PMID:11337471
A:Accession: C86850
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-627 <STO>
A:Cross-references: GB:AE005176; PID:gl2724829; PIDN:AAK05901.1; GSPDB:GN00146
A:Experimental source: strain IL1403
C:Genetics:
A:Gene: pepO

Query Match 71.4%; Score 30; DB 2; Length 627;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 RIKASKSF 9
|: ||| |||
Db 570 RMKASKEF 577

RESULT 15
A47098
lactococcal endopeptidase (EC 3.4.21.1) - Lactococcus lactis subsp. cremoris (strain P
C:Species: Lactococcus lactis subsp. cremoris
C:Date: 03-May-1994 #sequence_revision 03-May-1994 #text_change 15-Oct-1999
C:Accession: A47098
R:Mierau, I.; Tan, P.S.T.; Haandrikman, A.J.; Kok, J.; Leenhouts, K.J.; Konings, W.N.
J. Bacteriol. 175, 2087-2096, 1993
A:Title: Cloning and sequencing of the gene for a lactococcal endopeptidase, an enzym
A:Reference number: A47098; MUID:93209954
A:Accession: A47098
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-627 <MIE>
A:Cross-references: GB:L04938; NID:g293013; PIDN:AAA25204.1; PID:g293015
C:Genetics:
A:Gene: pepO
C:Keywords: hydrolase; oligopeptide transport

Query Match 71.4%; Score 30; DB 2; Length 627;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 2 RIKASKSF 9
|: ||| |||
Db 570 RMKASKEF 577

RESULT 16
F53290
endopeptidase PepO (EC 3.4.21.1) - Lactococcus lactis subsp. lactis
C:Species: Lactococcus lactis subsp. lactis
C:Date: 13-Sep-1995 #sequence_revision 13-Sep-1995 #text_change 15-Oct-1999
C:Accession: F53290
R:Tynkynen, S.; Buist, G.; Kunji, E.; Kok, J.; Poolman, B.; Venema, G.; Haandrikman,
J. Bacteriol. 175, 7523-7532, 1993

A:Title: Genetic and biochemical characterization of the oligopeptide transport system
A:Reference number: A53290; MUID:94064545
A:Accession: F53290
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-627 <TYN>
A:Cross-references: GB:L38760; NID:g308849; PIDN:AAA16168.1; PID:g308855
C:Genetics:
A:Gene: pepO
C:Keywords: hydrolase

Query Match 71.4%; Score 30; DB 2; Length 627;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKASKSF 9
||:|||||
Db 570 RMKASKEF 577

RESULT 17
T19678
hypothetical protein C33D9.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19678
R:Lloyd, C.
submitted to the EMBL Data Library, December 1995
A:Reference number: Z19162
A:Accession: T19678
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-875 <WIL>
A:Cross-references: EMBL:Z68159; PIDN:CAA92282.1; GSPDB:GN00022; CESP:C33D9.1
A:Experimental source: clone C33D9
C:Genetics:
A:Gene: CESP:C33D9.1
A:Map position: 4
A:Introns: 14/1; 74/3; 129/3; 171/3; 198/1; 241/2; 269/3; 352/3; 403/2; 515/3; 595/3; 65

Query Match 71.4%; Score 30; DB 2; Length 875;
Best Local Similarity 62.5%; Pred. No. 1.7e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 2 RIKASKSF 9
||:|||||
Db 233 RVKSKSY 240

RESULT 18
D82776
pyruvate dehydrogenase XF0669 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: D82776
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717
A:Note: for a complete list of authors see reference number A59328 below
A:Accession: D82776
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-985 <SIM>
A:Cross-references: GB:AE003911; GB:AE003849; NID:g9105548; PIDN:AAF83479.1; GSPDB:GN001
A:Experimental source: strain 9a5c
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrer, H
as-Neto, E.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laig
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuchako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: XF0669
C:Superfamily: pyruvate dehydrogenase (lipoamide); thiamin pyrophosphate-binding doma

Query Match 71.4%; Score 30; DB 2; Length 985;
Best Local Similarity 66.7%; Pred. No. 1.9e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
||:|||||
Db 560 RRTKSKSF 568

RESULT 19
H84545
probable ubiquitin-conjugating enzyme [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84545
R:Lip, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487
A:Accession: H84545
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1102 <STO>
A:Cross-references: GB:AE002093; NID:g3757521; PIDN:AAC64223.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g16920
A:Map position: 2

Query Match 71.4%; Score 30; DB 2; Length 1102;
Best Local Similarity 75.0%; Pred. No. 2.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
||:|||||
Db 257 QRKASSS 264

RESULT 20
S02041
dystrophin, muscle - chicken
N:Alternate names: duchenne muscular dystrophy protein
C:Species: Gallus gallus (chicken)
C:Date: 07-Sep-1990 #sequence_revision 27-Jun-1994 #text_change 16-Jul-1999
C:Accession: S02041; S02013; S71487
R:Lemaire, C.; Heilig, R.; Mandel, J.L.
Nucleic Acids Res. 16, 11815-11816, 1988
A:Title: Nucleotide sequence of chicken dystrophin cDNA.
A:Reference number: S02041; MUID:89098331
A:Accession: S02041
A:Status: translation not shown
A:Molecule type: mRNA
A:Residues: 1-3660 <LEM>
A:Cross-references: EMBL:X13369; NID:g63369; PIDN:CAA31746.1; PID:g63370
A:Note: 1869-His, 1885-Arg, and sequences lacking 1171-Met were also found
R:Lemaire, C.; Heilig, R.; Mandel, J.L.
EMBO J. 7, 4157-4162, 1988
A:Title: The chicken dystrophin cDNA: striking conservation of the C-terminal coding
A:Reference number: S02013; MUID:89210800
A:Accession: S02013
A:Status: nucleic acid sequence not shown

A: Experimental source: strain PCC 7120
C: Genetics:
A: Gene: asi3709

Query Match 69.0%; Score 29; DB 2; Length 65;
Best Local Similarity 55.6%; Pred. No. 20;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
| :|||
Db 37 QEVEANKSF 45

RESULT 22
B46322
16K protein - tobacco rattle virus (strain PLB)
C: Species: tobacco rattle virus, TRV
C: Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 23-Jul-1999
C: Accession: B46322
R: Angenent, G.C.; Posthumus, E.; Brederode, F.T.; Bol, J.F.
Virology 171, 271-274, 1989
A: Title: Genome structure of tobacco rattle virus strain PLB: further evidence on the
A: Reference number: A46322; MUID:89299469
A: Accession: B46322
A: Molecule type: genomic RNA
A: Residues: 1-141 <ANG>
A: Cross-references: GB:J04347; NID:g332219; PIDN:AAA47079.1; PID:g332221
C: Superfamily: tobacco rattle virus 16K protein

Query Match 69.0%; Score 29; DB 1; Length 141;
Best Local Similarity 55.6%; Pred. No. 45;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
| :|||
Db 90 ERIRAERSF 98

RESULT 23
H81251
NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain J Cj1570c [imported] - Campyloba
C: Species: Campylobacter jejuni
C: Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 15-Sep-2000
C: Accession: H81251
R: Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chli
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A: Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A: Reference number: A81250; MUID:20150912
A: Accession: H81251
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-172 <PAR>
A: Cross-references: GB:AL111168; NID:g9668971; PIDN:CAB73558.1; PID:g966
C: Experimental source: serotype O2, strain NCTC 11168
C: Genetics:
A: Gene: nuoJ; Cj1570c
C: Keywords: NAD; oxidoreductase

Query Match 69.0%; Score 29; DB 2; Length 172;
Best Local Similarity 55.6%; Pred. No. 55;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
| :|||
Db 84 EQLKAKSF 92

RESULT 24
C64176
hypothetical protein HI1715 - Haemophilus Influenzae (strain Rd KW20)
C: Species: Haemophilus influenzae
C: Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 11-Jan-2000

C:Accession: C64176
R: Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A: Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
A: Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A: Reference number: A64000; MUID: 95350630
A: Accession: C64176
A: Status: nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-182 <YIGR>
A: Cross-references: GB:U32844; GB:L42023; NID:gl574563; PIDN:AAC23360.1; PID:gl574570; T
C: Note: best homolog was a hypothetical protein from Escherichia coli
C: Superfamily: human conserved hypothetical protein DKFp566E144.1

Query Match 69.0%; Score 29; DB 2; Length 182;
Best Local Similarity 85.7%; Pred. No. 56;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASK 7
:|||||
Db 73 ERIKASK 79
:|||||

RESULT 25
C90678
thiogalactoside acetyltransferase [imported] - Escherichia coli (strain O157:H7, substra
C: Species: Escherichia coli
C: Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Aug-2001
C: Accession: C90678
R: Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A: Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genc
A: Reference number: A9629; MUID:21156231; PMID:11258796
A: Accession: C90678
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-203 <HAY>
A: Cross-references: GB:BA000007; PIDN:BA833818.1; PID:gl3359852; GSPDB:GN00154
A: Experimental source: strain O157:H7, substrain RMD 0509952
C: Genetics:
A: Gene: EC0395
C: Superfamily: galactoside acetyltransferase

Query Match 69.0%; Score 29; DB 2; Length 203;
Best Local Similarity 66.7%; Pred. No. 65;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
:|||||
Db 7 ERIKAGKLF 15
:|||||

RESULT 26
G85528
thiogalactoside acetyltransferase [imported] - Escherichia coli (strain O157:H7, substra
C: Species: Escherichia coli
C: Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C: Accession: G85528
R: Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
iller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A: Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A: Reference number: A85480; MUID:21074935; PMID:11206551
A: Accession: G85528
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-203 <STO>
A: Cross-references: GB:AE005174; NID:gl12513173; PIDN:AAG54691.1; GSPDB:GN00145; UWGP:204
A: Experimental source: strain O157:H7, substrain EDL933
C: Genetics:

A:Gene: lacA
C:Superfamily: galactoside acetyltransferase

Query Match 69.0%; Score 29; DB 2; Length 203;
Best Local Similarity 66.7%; Pred. No. 65;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
:|||||
Db 7 ERIKAGKLF 15
:|||||

RESULT 27
E84339
quinolinate phosphoribosyltransferase [imported] - Halobacterium sp. NRC-1
C: Species: Halobacterium sp. NRC-1
C: Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C: Accession: E84339
R: Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky
; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Ja
Jung, K.H.; Alam, M.; Freitas, T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A: Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.,
A: Title: Genome sequence of Halobacterium species NRC-1.
A: Reference number: A84160; MUID:20504483
A: Accession: E84339
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-268 <STO>
A: Cross-references: GB:AE004437; NID:gl0581326; PIDN:AAG20081.1; GSPDB:GN00138
C: Genetics:
A: Gene: nadC
C: Superfamily: nicotinate-nucleotide pyrophosphorylase (carboxylating)

Query Match 69.0%; Score 29; DB 2; Length 268;
Best Local Similarity 55.6%; Pred. No. 86;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
:|||||
Db 172 ERFRAAKSF 180
:|||||

RESULT 28
AE2568
WD repeat protein [imported] - Anabaena sp. (strain PCC 7120) plasmid pCC7120delta
C: Species: Anabaena sp.
A: Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C: Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C: Accession: AE2568
R: Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriqu
Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata
DNA Res. 8, 205-213, 2001
A: Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium
A: Reference number: AB1807; MUID:21595285; PMID:11759840
A: Accession: AE2568
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-349 <KUR>
A: Cross-references: GB:AP003604; PIDN:BA877463.1; PID:gl17134907; GSPDB:GN00183
A: Experimental source: strain PCC 7120
C: Genetics:
A: Gene: all8544
A: Genome: Plasmid

Query Match 69.0%; Score 29; DB 2; Length 349;
Best Local Similarity 85.7%; Pred. No. 1.1e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IKASKSF 9
:|||||
Db 267 IKANKSF 273
:|||||

RESULT 29

T22587

hypothetical protein F53F8.5 - *Caenorhabditis elegans*C:Species: *Caenorhabditis elegans*

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T22587

R:Matthews, L.

submitted to the EMBL Data Library, November 1996

A:Reference number: Z19585

A:Accession: T22587

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-368 <WIL>

A:Cross-references: EMBL:Z81547; PIDN:CAB04458.1; GSPDB:GN00023; CESP:F53F8.5

A:Experimental source: clone F53F8

C:Genetics:

A:Gene: CESP:F53F8.5

A:Map position: 5

A:Introns: 54/73; 146/1; 205/2; 284/3

Query Match 69.0%; Score 29; DB 2; Length 368;

Best Local Similarity 75.0%; Pred. No. 1.2e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKS 8

DB 59 RRIKAASKS 66

RESULT 30

G97300

asparaginyl-tRNA synthetase [imported] - *Clostridium acetobutylicum*C:Species: *Clostridium acetobutylicum*

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 30-Sep-2001

C:Accession: G97300

R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: G97300

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-463 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK81194.1; PID:g15026334; GSPDB:GN00168

A:Experimental source: *Clostridium acetobutylicum* ATCC824

C:Genetics:

A:Gene: CAC3260

C:Superfamily: lysine--tRNA ligase

Query Match 69.0%; Score 29; DB 2; Length 463;

Best Local Similarity 85.7%; Pred. No. 1.5e+02;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IKASKSF 9

DB 30 IRASKSF 36

RESULT 31

JC5282

phosphogluconate dehydrogenase (decarboxylating) (EC 1.1.1.44) - *Actinobacillus actinomy-*

N:Alternate names: 6-Phosphogluconic carboxylase; 6-Phosphogluconic dehydrogenase; Phospho-

C:Species: *Actinobacillus actinomycescomitans*

C:Date: 16-Apr-1997 #sequence_revision 09-May-1997 #text_change 20-Jun-2000

C:Accession: JC5282

R:Yoshida, Y.; Nakano, Y.; Yamashita, Y.; Koga, T.

Biochem. Biophys. Res. Commun. 230, 220-225, 1997

A:Title: The gnd gene encoding a novel 6-phosphogluconate dehydrogenase and its adjacent

A:Reference number: JC5282; MUID:97148607

A:Accession: JC5282

A:Molecule type: DNA

A:Residues: 1-484 <YOS>

A:Cross-references: DBJ:D88189; NID:g1651206; PIDN:BAA13558.1; PID:g1651212

A:Experimental source: strain Y4

C:Comment: This enzyme is involved in the pentose phosphate pathway. It plays roles in wall constituents, and production of NADPH for reductive biosynthesis.

C:Genetics:

A:Gene: gnd

C:Superfamily: phosphogluconate dehydrogenase (decarboxylating); 3-hydroxyisobutyrate

C:Keywords: Oxidoreductase

F:7-287/Domain: 3-hydroxyisobutyrate dehydrogenase homology <HIB>

Query Match 69.0%; Score 29; DB 2; Length 484;

Best Local Similarity 66.7%; Pred. No. 1.6e+02;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

DB 297 ORVAASKLF 305

RESULT 32

G71081

probable helicase protein PH0917 - *Pyrococcus horikoshii*C:Species: *Pyrococcus horikoshii*

C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 19-Jan-2001

C:Accession: G71081

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic

A:Reference number: A71000; MUID:98344137

A:Accession: G71081

A:Status: preliminary; nucleic-acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-526 <KAW>

A:Cross-references: GB:AP000004; NID:g3236131; PIDN:BAA30013.1; PID:g3257330

A:Experimental source: strain OT3

A:Note: this accession replaces an interim accession for a sequence replaced by GenBa

C:Genetics:

A:Gene: PH0917

C:Superfamily: *Pyrococcus horikoshii* probable helicase PH0917

C:Keywords: ATP; nucleotide binding; P-loop

F:54-61/Region: nucleotide-binding motif A (P-loop)

F:165-170/Region: nucleotide-binding motif B

F:169-172/Region: DEAH motif

Query Match 69.0%; Score 29; DB 1; Length 526;

Best Local Similarity 75.0%; Pred. No. 1.7e+02;

Matches 6; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2 RIKASKSF 9

DB 144 RIKAKREF 151

RESULT 33

B82883

hypothetical protein UU499 [imported] - *Ureaplasma urealyticum*C:Species: *Ureaplasma urealyticum*

C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: B82883

R:Glass, J.I.; Lefkowitz, E.J.; Glass, J.S.; Heiner, C.R.; Chen, E.Y.; Cassell, G.H.

submitted to GenBank, February 2000

A:Description: The complete sequence of *Ureaplasma urealyticum*: Alternate views of a

A:Reference number: A82870

A:Accession: B82883

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-563 <GLA>

A:Cross-references: GB:AE002148; GB:AF222894; NID:g6899495; PIDN:AAF30911.1; GSPDB:GN

A:Experimental source: serovar 3; biovar 1

C:Genetics:

A:Gene: UU499

A:Genetic code: SGC3

Query Match 69.0%; Score 29; DB 2; Length 563;
Best Local Similarity 66.7%; Pred. No. 1.8e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

|||||

Db 459 QKINASNSF 467

RESULT 34

T48273

hypothetical protein T22P11.110 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000

C:Accession: T48273

R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancroft,

submitted to the Protein Sequence Database, April 2000

A:Reference number: Z24490

A:Accession: T48273

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-584 <BEV>

A:Cross-references: EMBL:AL162971

A:Experimental source: cultivar Columbia; BAC clone T22P11

C:Genetics:

A:Map position: 5

A:Introns: 18/3; 48/2; 96/3; 136/1; 517/3; 545/1; 564/3

A:Note: T22P11.110

Query Match

Best Local Similarity 69.0%; Score 29; DB 2; Length 584;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKASKSF 9

|||||

Db 456 RRVKSKAF 463

RESULT 35

I37216

calicin - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 21-Jul-2000

C:Accession: I37216; S55466

R:Von Bulow, M.; Heid, H.; Hess, H.; Franke, W.W.

Exp. Cell Res. 219, 407-413, 1995

A:Title: Molecular nature of calicin, a major basic protein of the mammalian sperm head

A:Reference number: I37216; MUID:95369354

A:Accession: I37216

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-585 <RES>

A:Cross-references: EMBL:Z46967; NID:g854099; PIDN:CAA87088.1; PID:g854100

Query Match

Best Local Similarity 69.0%; Score 29; DB 2; Length 585;

Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKS 8

|||||

Db 490 QKVKASKT 497

RESULT 36

AB2028

hypothetical protein all1776 [imported] - Anabaena sp. (strain PCC 7120)

C:Species: Anabaena sp.

A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002

C:Accession: AB2028

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi,

Nakazaki, N.; Shimpou, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata:
DNA Res. 8, 205-213, 2001

A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AB2028

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-669 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA873475.1; PID:g17130866; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all1776

Query Match

Best Local Similarity 69.0%; Score 29; DB 2; Length 669;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

|||||

Db 431 QGIKATKSY 439

RESULT 37

T46476

hypothetical protein DKFZp434C0931.1 - human

C:Species: Homo sapiens (man)

C:Date: 04-Feb-2000 #sequence_revision 04-Feb-2000 #text_change 04-Feb-2000

C:Accession: T46476

R:Blum, H.; Bauersachs, S.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.

submitted to the Protein Sequence Database, January 2000

A:Reference number: Z23034

A:Accession: T46476

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-691 <AAA>

A:Cross-references: EMBL:AL137718

A:Experimental source: adult testis; clone DKFZp434C0931

C:Genetics:

A:Note: DKFZp434C0931.1

Query Match

Best Local Similarity 69.0%; Score 29; DB 2; Length 691;

Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

|||||

Db 565 EEIKRKSXF 573

RESULT 38

AC1419

DNA topoisomerase III homolog topB [imported] - Listeria monocytogenes (strain EGD-e)

C:Species: Listeria monocytogenes

C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C:Accession: AC1419

R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Anand, A.; Baquero, F.; Berche, P.; Bloer,

D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entian, K.D.; Fsihl,

Science 294, 849-852, 2001

A:Authors: Krefte, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.;

ok, C.; Schluter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehl-

A:Title: Comparative genomics of Listeria species

A:Reference number: AB1077; MUID:21537279; PMID:11679669

A:Accession: AC1419

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-717 <GLA>

A:Cross-references: GB:NC_003210; PIDN:CAD00969.1; PID:g16412256; GSPDB:GN00177

A:Experimental source: strain EGD-e

C:Genetics:

A:Gene: topB

C:Superfamily: DNA topoisomerase I topA

Query Match 69.0%; Score 29; DB 2; Length 717;
Best Local Similarity 85.7%; Pred. No. 2.3e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IKASKSF 9
|||||

DB 345 IKANKSF 351

RESULT 39

E86294

hypothetical protein T24D18.9 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001

C:Accession: E86294

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: E86294

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-738 <STO>

A:Cross-references: GB:AE005172; NID:g6587805; PIDN:AAF18496.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 69.0%; Score 29; DB 2; Length 738;
Best Local Similarity 86.7%; Pred. No. 2.4e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKF 9
:::|||||

DB 47 EKKKASKF 55

RESULT 40

C86168

hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C:Accession: C86168

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719

A:Accession: C86168

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-743 <STO>

A:Cross-references: GB:AE005172; NID:g4204309; PIDN:AAD10690.1; GSPDB:GN00141

C:Genetics:

A:Map position: 1

Query Match 69.0%; Score 29; DB 2; Length 743;
Best Local Similarity 75.0%; Pred. No. 2.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKASKSF 9

DB 46 RIKARRSF 53
|||||

RESULT 41

S67164

probable membrane protein YOR267c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein O5420

C:Species: Saccharomyces cerevisiae

C>Date: 12-Jul-1996 #sequence_revision 12-Jul-1996 #text_change 24-Sep-1999

C:Accession: S67164; S67169; S72039

R:Jauniaux, J.C.; Polirey, R.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67143

A:Accession: S67164

A:Molecule type: DNA

A:Residues: 1-759 <JAU>

A:Cross-references: EMBL:Z75175; NID:gl420599; PIDN:CAA99490.1; PID:e252116; PID:gl42

A:Experimental source: strain S288C

R:Cheret, G.; Sor, F.

submitted to the Protein Sequence Database, July 1996

A:Reference number: S67169

A:Accession: S67169

A:Molecule type: DNA

A:Residues: 1-367 <CHE>

A:Cross-references: EMBL:Z75175; MIPS:YOR267c

A:Experimental source: strain S288C

R:Cheret, G.; Bernardi, A.; Sor, F.

Yeast 12, 1059-1064, 1996

A>Title: DNA sequence analysis of the VP1-SNF2 region on chromosome XV of Saccharomy

A:Reference number: S72039; MUID:97051594

A:Accession: S72039

A>Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-367 <CHW>

A:Cross-references: EMBL:X89633

A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1995

C:Genetics:

A:Map position: 15R

A>Note: YOR267c

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom

C:Keywords: ATP; transmembrane protein

F:213-496/Domain: protein kinase homology <KIN>

F:221-229/Region: protein kinase ATP-binding motif

F:400-416/Domain: transmembrane #status predicted <TMM>

F:628-656/Region: glutamine-rich

Query Match 69.0%; Score 29; DB 2; Length 759;
Best Local Similarity 66.7%; Pred. No. 2.5e+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
|:::|||||

DB 30 QLIRSKSF 38

RESULT 42

A40985

projectin - fruit fly (Drosophila melanogaster) (fragments)

C:Species: Drosophila melanogaster

C>Date: 28-May-1992 #sequence_revision 28-May-1992 #text_change 16-Feb-1997

C:Accession: A40985

R:Ayme-Southgate, A.; Vigoreaux, J.; Benlan, G.; Pardue, M.L.

Proc. Natl. Acad. Sci. U.S.A. 88, 7973-7977, 1991

A>Title: Drosophila has a twitchin/titin-related gene that appears to encode projecti

A:Reference number: A40985; MUID:91376068

A:Accession: A40985

A>Status: preliminary

A:Molecule type: nucleic acid sequence not shown; not compared with conceptual

A:Residues: 1-940 <AYW>

A:Cross-references: GB:M73433

C:Genetics:

A:Gene: FlyBase:bt

A;Cross-references: FlyBase:FBgn0005666

C;Superfamily: fibronectin type III repeat homology

Query Match 69.0%; Score 29; DB 2; Length 940;
Best Local Similarity 75.0%; Pred. No. 3.1e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIRAKSF 9

||||:|

Db 651 RIRAGQSF 658

RESULT 43

T15297

hypothetical protein B0280.2 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 04-Mar-2000

C;Accession: T15297

R;Fulton, L.

submitted to the EMBL Data Library, June 1994

A;Description: The sequence of C. elegans cosmid B0280.

A;Reference number: S48966

A;Accession: T15297

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-1018 <FUL>

A;Cross-references: EMBL:U10438; NID:g500762; PID:g500764; PIDN:AA19081.1; CESP:B0280.2

A;Experimental source: strain Bristol N2

C;Genetics:

A;Gene: CESP:B0280.2

A;Introns: 27/3; 134/1; 196/3; 245/3; 312/1; 445/3; 546/3; 585/3; 616/3; 647

C;Superfamily: Caenorhabditis elegans hypothetical protein B0280.2

Query Match

Best Local Similarity 69.0%; Score 29; DB 2; Length 1018;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIRAKSF 9

::| ||||

Db 240 KVTKSF 247

RESULT 44

S24600

projectin - fruit fly (Drosophila melanogaster)

C;Species: Drosophila melanogaster

C;Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 24-Sep-1998

C;Accession: S24600

R;Fyrberg, E.

submitted to the EMBL Data Library, May 1992

A;Reference number: S24600

A;Accession: S24600

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-1742 <FYR>

A;Cross-references: EMBL:X66018; NID:g8377; PID:g8378

C;Genetics:

A;Gene: FlyBase:bt

A;Cross-references: FlyBase:FBgn0005666

Query Match

Best Local Similarity 69.0%; Score 29; DB 2; Length 1742;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIRAKSF 9

||||:|

Db 630 RIRAGQSF 637

RESULT 45

T13931

projectin - fruit fly (Drosophila melanogaster) (fragment)

C;Species: Drosophila melanogaster

C;Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000

C;Accession: T13931

R;Daley, J.; Southgate, R.; Ayme-Southgate, A.

J. Mol. Biol. 279, 201-210, 1998

A;Title: Structure of the Drosophila projectin protein: isoforms and implication for

A;Reference number: Z17815; MUID:98300339

A;Accession: T13931

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-6658 <DAL>

A;Cross-references: EMBL:AF047475; NID:g3337430; PID:g3337431; PIDN:AAC27550.1

C;Genetics:

A;Gene: projectin

A;Cross-references: FlyBase:FBgn0005666

A;Map position: 4

A;Note: intron positions not resolved (incomplete sequence)

C;Keywords: muscle

Query Match

Best Local Similarity 69.0%; Score 29; DB 2; Length 6658;

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIRAKSF 9

||||:|

Db 2718 RIRAGQSF 2725

Search completed: November 6, 2002, 12:08:25

Job time : 14 secs

THIS PAGE BLANK (USP 12)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 8.11111 Seconds
(without alignments)
27.102 Million cell updates/sec

Title: US-09-905-083-99
Perfect score: 42
Sequence: 1 QRIKASKSF 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match %	Length	DB ID	Description
1	42	100.0	9	4	US-09-502-600-99
2	42	100.0	9	4	US-09-502-600-130
3	42	100.0	154	4	US-09-261-416-7
4	42	100.0	224	4	US-08-944-483-33
5	42	100.0	225	2	US-08-557-146-12
6	42	100.0	225	2	US-09-027-337-4
7	42	100.0	225	2	US-09-154-344-12
8	42	100.0	253	2	US-08-557-146-2
9	42	100.0	253	2	US-08-824-874-3
10	42	100.0	253	2	US-09-154-344-2
11	42	100.0	253	3	US-08-930-188-2
12	42	100.0	253	4	US-09-210-084-3
13	42	100.0	253	5	PCT-US96-04294-2
14	36	85.7	9	4	US-09-502-600-97
15	31	73.8	250	1	US-07-914-282D-6
16	31	73.8	250	1	US-08-276-887A-6
17	31	73.8	250	5	PCT-US93-02460-6
18	31	73.8	2860	2	US-08-826-267-2
19	30	71.4	666	4	US-08-982-785A-11
20	28	66.7	61	1	US-08-194-211A-3
21	28	66.7	61	4	US-08-456-748B-3
22	28	66.7	500	4	US-09-442-100-14
23	28	66.7	1824	2	US-08-680-327-3
24	28	66.7	1824	4	US-09-228-246-2
25	27	64.3	9	4	US-09-502-600-106
26	27	64.3	99	3	US-09-173-581-8
27	27	64.3	99	4	US-09-420-915-8

28	27	64.3	100	3	US-09-034-916-11	Sequence 11, Appl
29	27	64.3	147	6	5175147-8	Patent No. 5175147
30	27	64.3	202	6	5189147-6	Patent No. 5189147
31	27	64.3	325	2	US-09-018-576-3	Sequence 3, Appl1
32	27	64.3	325	2	US-09-018-576-12	Sequence 12, Appl
33	27	64.3	325	3	US-09-248-137-3	Sequence 3, Appl1
34	27	64.3	325	3	US-09-248-137-12	Sequence 12, Appl
35	27	64.3	325	4	US-09-000-016-7	Sequence 7, Appl1
36	27	64.3	734	4	US-09-000-016-4	Sequence 4, Appl1
37	27	64.3	823	4	US-09-000-016-2	Sequence 2, Appl1
38	26	61.9	65	1	US-08-691-641-11	Sequence 11, Appl
39	26	61.9	68	1	US-08-606-789-8	Sequence 8, Appl1
40	26	61.9	68	2	US-09-111-348-8	Sequence 8, Appl1
41	26	61.9	69	5	PCT-US95-06406A-6	Sequence 6, Appl1
42	26	61.9	158	2	US-08-917-456-2	Sequence 2, Appl1
43	26	61.9	158	4	US-09-229-804-2	Sequence 2, Appl1
44	26	61.9	194	3	US-09-032-372-13	Sequence 13, Appl
45	26	61.9	194	3	US-09-078-317-12	Sequence 12, Appl

ALIGNMENTS

RESULT 1
US-09-502-600-99
; Sequence 99, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 99
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 91-99 of the SCCE protein
US-09-502-600-99

Query Match 100.0%; Score 42; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 1 QRIKASKSF 9

RESULT 2
US-09-502-600-130
; Sequence 130, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; FILE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 130
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:

```
; OTHER INFORMATION: Residues 91-99 of the SCCE protein
US-09-502-600-130
Query Match      100.0%; Score 42; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
    |
Db 1 QRIKASKSF 9
    |

RESULT 3
US-09-261-416-7
; Sequence 7, Application US/09261416A
; Patent No. 6291663
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Underwood, Lowell J.
; TITLE OF INVENTION: TADG-12: A No. 6291663el Transmembrane Serine Protease
; TITLE OF INVENTION: Overexpressed in Ovarian Carcinoma
; FILE REFERENCE: D6192
; CURRENT APPLICATION NUMBER: US/09/261.416A
; CURRENT FILING DATE: 1999-03-03
; NUMBER OF SEQ ID NOS: 14
; SEQ ID NO 7
; LENGTH: 154
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of stratum corneum
; OTHER INFORMATION: chymotryptic enzyme (Scce) homologous to similar
; OTHER INFORMATION: domain in TADG-12
US-09-261-416-7

Query Match      100.0%; Score 42; DB 4; Length 154;
Best Local Similarity 100.0%; Pred. No. 0.095;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
    |
Db 37 QRIKASKSF 45
    |

RESULT 4
US-08-944-483-33
; Sequence 33, Application US/08944483
; Patent No. 6232456
; GENERAL INFORMATION:
; APPLICANT: COHEN, MAURICE
; APPLICANT: COLPITTS, TRACEY L.
; APPLICANT: FRIEDMAN, PAULA N.
; APPLICANT: GRANADOS, EDWARD N.
; APPLICANT: KLASS, MICHAEL R.
; APPLICANT: RUSSELL, JOHN C.
; APPLICANT: STEWART, KENT D.
; APPLICANT: STROUPE, STEVEN D.
; TITLE OF INVENTION: NOVEL SERINE PROTEASE REAGENTS
; TITLE OF INVENTION: AND METHODS USEFUL FOR DETECTING AND TREATING DISEASES
; TITLE OF INVENTION: OF THE PROSTATE
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: IL
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/944.483
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Becker, Cheryl L.
; REGISTRATION NUMBER: 35,441
; REFERENCE/DOCKET NUMBER: 6183-US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 847/935-1729
; TELEFAX: 847/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 224 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: No. 6232456e
US-08-944-483-33

Query Match      100.0%; Score 42; DB 4; Length 224;
Best Local Similarity 100.0%; Pred. No. 0.14;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
    |
Db 62 QRIKASKSF 70
    |

RESULT 5
US-08-557-146-12
; Sequence 12, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
```

US-08-557-146-12

Query Match 100.0%; Score 42; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.14; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
|||||
Db 63 QRIKASKSF 71

RESULT 6

US-09-027-337-4
; Sequence 4, Application US/09027337B
; Patent No. 5972616
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; APPLICANT: Tanimoto, Hiroto
; TITLE OF INVENTION: TADG-15: An Extracellular Serine Protease Overexpressed in
; TITLE OF INVENTION: TADG-15: Breast and Ovarian Carcinomas
; FILE REFERENCE: D6064
; CURRENT APPLICATION NUMBER: US/09/027,337B
; CURRENT FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 4
; LENGTH: 225
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Serine protease catalytic domain of Scce homologous to
; OTHER INFORMATION: similar domain in TADG-15
US-09-027-337-4

Query Match 100.0%; Score 42; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.14; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
|||||
Db 63 QRIKASKSF 71

RESULT 7

US-09-154-344-12
; Sequence 12, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:

; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 225 amino acids
; TYPE: amino acids
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: polypeptide
US-09-154-344-12

Query Match 100.0%; Score 42; DB 2; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.14; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
|||||
Db 63 QRIKASKSF 71

RESULT 8

US-08-557-146-2
; Sequence 2, Application US/08557146
; Patent No. 5834290
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; TITLE OF INVENTION: Enzyme (SCCE)
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-557-146-2

Query Match 100.0%; Score 42; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.16; 0; Indels 0; Gaps 0;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
|||||
Db 91 QRIKASKSF 99

RESULT 9

US-08-824-874-3
; Sequence 3, Application US/08824874
; Patent No. 5962300
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; TITLE OF INVENTION: NOVEL KALLIKREIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/824,874
; FILING DATE: Filed Herewith
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0252 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-855-0555
; TELEFAX: 415-845-4166
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 532504
US-08-824-874-3

Query Match 100.0%; Score 42; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ORIKASKSF 9
Db 91 ORIKASKSF 99

RESULT 10

US-09-154-344-2
; Sequence 2, Application US/09154344
; Patent No. 5981256
; GENERAL INFORMATION:
; APPLICANT: Egelrud, Torbjorn
; APPLICANT: Hansson, Lennart
; TITLE OF INVENTION: Recombinant Stratum Corneum Chymotryptic
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case, Patent Department
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10036-2787

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/154,344
; FILING DATE: 16-SEP-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/557,146
; FILING DATE: 14-DEC-1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner, Richard J.
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-181
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 819-8783
; TELEFAX: (212) 354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 253 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-154-344-2

Query Match 100.0%; Score 42; DB 2; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 ORIKASKSF 9
Db 91 ORIKASKSF 99

RESULT 11

US-08-930-188-2
; Sequence 2, Application US/08930188
; Patent No. 6093397
; GENERAL INFORMATION:
; APPLICANT: Dixon, Eric P.
; APPLICANT: Johnstone, Edward M.
; APPLICANT: Little, Sheila P.
; TITLE OF INVENTION: ANGIOID PRECURSOR PROTEIN PROTEASE AND
; RELATED NUCLEIC ACIDS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: United States of America
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/930,188
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/416,257
; FILING DATE: 04-APR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Blalock, Donna K.
; REGISTRATION NUMBER: 38,082
; REFERENCE/DOCKET NUMBER: X9239
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-277-1090

TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-930-188-2

Query Match 100.0%; Score 42; DB 3; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9
| | | | | | | | | |
DB 91 ORIKASKSF 99

RESULT 12
US-09-210-084-3
Sequence 3, Application US/09210084
Patent No. 6197511
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Lal, Preeti
TITLE OF INVENTION: NOVEL KALLIKREIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/210,084
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/824,874
FILING DATE:

ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0252 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-853-0555
TELEFAX: 415-845-4166

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 532504
US-09-210-084-3

Query Match 100.0%; Score 42; DB 4; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9
| | | | | | | | | |
DB 91 ORIKASKSF 99

RESULT 13
PCT-US96-04294-2
Sequence 2, Application PC/TUS9604294
GENERAL INFORMATION:
APPLICANT: Dixon, Eric P.
APPLICANT: Johnstone, Edward M.
APPLICANT: Little, Sheila P.
TITLE OF INVENTION: AMYLOID PRECURSOR PROTEIN PROTEASE AND
RELATED NUCLEIC ACIDS
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Eli Lilly and Company
STREET: Lilly Corporate Center
CITY: Indianapolis
STATE: Indiana
COUNTRY: United States of America
ZIP: 46285
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/04294
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/416,257
FILING DATE: 04-APR-1995
ATTORNEY/AGENT INFORMATION:
NAME: Blalock, Donna K.
REGISTRATION NUMBER: 38,082
REFERENCE/DOCKET NUMBER: X9239
TELECOMMUNICATION INFORMATION:
TELEPHONE: 317-277-1090
TELEFAX: 317-276-3861
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 253 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-04294-2

Query Match 100.0%; Score 42; DB 5; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.16;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9
| | | | | | | | | |
DB 91 ORIKASKSF 99

RESULT 14
US-09-502-600-97
Sequence 97, Application US/09502600A
Patent No. 6294344
GENERAL INFORMATION:
APPLICANT: O'Brien, Timothy J.
TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
Ovarian Cancer
FILE REFERENCE: D6223CIP-C
CURRENT FILING DATE: 2000-02-11
CURRENT APPLICATION NUMBER: US/09/502,600A
PRIOR APPLICATION NUMBER: 09/039,211
PRIOR FILING DATE: 03-14-1998
NUMBER OF SEQ ID NOS: 136
SEQ ID NO 97
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens

; FEATURE:
; OTHER INFORMATION: Residues 90-98 of the SCE protein
US-09-502-600-97

Query Match 85.7%; Score 36; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 QRIKASKS 8
|:|:|:|:|
Db 2 QRIKASKS 9

RESULT 15
US-07-914-282D-6
; Sequence 6, Application US/07914282D
; Patent No. 5364787
; GENERAL INFORMATION:
; APPLICANT: Orser, Cindy S. and Xun, Luying
; TITLE OF INVENTION: GENES AND ENZYMES INVOLVED IN
; TITLE OF INVENTION: THE MICROBIAL DEGRADATION OF
; TITLE OF INVENTION: PENTACHLOROPHENOL
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ramon A. Klitzke II
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1/PC Gene
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/914,282D
; FILING DATE: July 13, 1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/856,015
; FILING DATE: March 23, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ramon A. Klitzke II
; REGISTRATION NUMBER: 30,188
; REFERENCE/DOCKET NUMBER: 2815-36746
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 250 amino acid residues
; TYPE: Amino Acid
; STRANDEDNESS:
; TOPOLOGY: Linear
; MOLECULE TYPE: Protein
; DESCRIPTION: Theoretical translation of open reading frame of pcpC gene
; HYPOTHETICAL: Yes
; ANTI-SENSE: No
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Flavobacterium sp. Strain ATCC 39723
US-07-914-282D-6

Query Match 73.8%; Score 31; DB 1; Length 250;
Best Local Similarity 66.7%; Pred. No. 31;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKS 9
|:|:|:|:|
Db 229 QRMKARRSF 237

RESULT 16
US-08-276-887A-6
; Sequence 6, Application US/08276887A
; Patent No. 5512478
; GENERAL INFORMATION:
; APPLICANT: Orser, Cindy S. and Xun, Luying
; TITLE OF INVENTION: GENES AND ENZYMES INVOLVED
; TITLE OF INVENTION: IN THE MICROBIAL
; TITLE OF INVENTION: DEGRADATION OF
; TITLE OF INVENTION: PENTACHLOROPHENOL
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ramon A. Klitzke II
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: WordPerfect 5.1/PC Gene
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/276,887A
; FILING DATE: July 18, 1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/914,282
; FILING DATE: July 13, 1992
; APPLICATION NUMBER: 07/856,015
; FILING DATE: March 23, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Ramon A. Klitzke II
; REGISTRATION NUMBER: 30,188
; REFERENCE/DOCKET NUMBER: 2815-36746
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 250 amino acid residues
; TYPE: Amino Acid
; STRANDEDNESS:
; TOPOLOGY: Linear
; MOLECULE TYPE: Protein
; DESCRIPTION: Theoretical translation of open reading frame of pcpC gene
; HYPOTHETICAL: Yes
; ANTI-SENSE: No
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; ORGANISM: Flavobacterium sp. Strain
; ORGANISM: ATCC 39723
US-08-276-887A-6

Query Match 73.8%; Score 31; DB 1; Length 250;
Best Local Similarity 66.7%; Pred. No. 31;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9
|:|:|:|:|
Db 229 QRMKARRSF 237

RESULT 17
PCT-US93-02460-6
; Sequence 6, Application PC/TUS9302460
; GENERAL INFORMATION:
; APPLICANT: Orser, Cindy S. and Xun, Luying

```
/
/ TITLE OF INVENTION: GENES AND ENZYMES INVOLVED IN THE
/ TITLE OF INVENTION: MICROBIAL DEGRADATION OF PENTACHLOROPHENOL
/ NUMBER OF SEQUENCES: 6
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: William D. Noonan, M.D.
/ STREET: One World Trade Center
/ STREET: 121 S.W. Salmon Street
/ STREET: Suite 1600
/ CITY: Portland
/ STATE: Oregon
/ COUNTRY: United States of America
/ ZIP: 97204
/
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette, 3.5-inch
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: MS DOS
/ SOFTWARE: WordPerfect 5.1/PC Gene
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US93/02460
/ FILING DATE: 19930319
/ CLASSIFICATION:
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/07/856,015
/ FILING DATE: March 23, 1992
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/07/914,282
/ FILING DATE: July 13, 1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: William D. Noonan, M.D.
/ REGISTRATION NUMBER: 30878
/ REFERENCE/DOCKET NUMBER: 2815-36746-WDN
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (503) 226-7391
/ TELEFAX: (503) 228-9446
/ INFORMATION FOR SEQ ID NO: 6:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 250 amino acids
/ TYPE: AMINO ACID
/ STRANDEDNESS:
/ TOPOLOGY: Linear
/ MOLECULE TYPE: Protein
/ DESCRIPTION: Theoretical translation of open reading
/ DESCRIPTION: frame of pcpc gene
/ HYPOTHETICAL: Yes
/ ANTI-SENSE: NO
/ FRAGMENT TYPE:
/ ORIGINAL SOURCE:
/ ORGANISM: Flavobacterium sp. Strain ATCC 39723
/ PCT-US93-02460-6
/
/ Query Match 73.8%; Score 31; DB 5; Length 250;
/ Best Local Similarity 66.7%; Pred. No. 31;
/ Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
/
/ QY 1 QRIKASKSF 9
/ Db 229 QRMKARSF 237
/
/ RESULT 18
/ US-08-826-267-2
/ Sequence 2, Application US/08826267
/ Patent No. 5994070
/ GENERAL INFORMATION:
/ APPLICANT: Streuli, Michel
/ TITLE OF INVENTION: No. 5994070el TRIO Molecules and Uses Related Thereto
/ NUMBER OF SEQUENCES: 2
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: LAHIVE & COCKFIELD
/ STREET: 28 State Street
/ CITY: Boston
/ STATE: Massachusetts
/ COUNTRY: USA
/
/ Query Match 73.8%; Score 31; DB 2; Length 2860;
/ Best Local Similarity 66.7%; Pred. No. 3.5e+02;
/ Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 QRIKASKSF 9
/ Db 225 QRIQSSESF 233
/
/ RESULT 19
/ US-08-982-785A-11
/ Sequence 11, Application US/08982785A
/ Patent No. 6258929
/ GENERAL INFORMATION:
/ APPLICANT: Kosik, Kenneth S.
/ APPLICANT: Zhou, Jianhua
/ TITLE OF INVENTION: ALARM RELATED PEPTIDES AND
/ TITLE OF INVENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM
/ NUMBER OF SEQUENCES: 13
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Fish & Richardson P.C.
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: FastSeq for Windows Version 2.0b
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/982,785A
/ FILING DATE: 02-DEC-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/031,556
/ FILING DATE: 02-DEC-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Freeman, John W.
/ REGISTRATION NUMBER: 29,066
/ REFERENCE/DOCKET NUMBER: 05311/018001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617/542-5070
/ TELEFAX: 617/542-8906
/ TELEX: 200154
```

```
/
/ ZIP: 02109-1875
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/826,267
/ FILING DATE: 1997
/ CLASSIFICATION: 800
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: 60/014,214
/ FILING DATE: 27 MARCH (1996)
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Amy E. Mandragouras
/ REGISTRATION NUMBER: 36,207
/ REFERENCE/DOCKET NUMBER: DFN-010
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (617)227-7400
/ TELEFAX: (617)227-5941
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 2860 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-826-267-2
/
/ Query Match 73.8%; Score 31; DB 2; Length 2860;
/ Best Local Similarity 66.7%; Pred. No. 3.5e+02;
/ Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
/
/ QY 1 QRIKASKSF 9
/ Db 225 QRIQSSESF 233
/
/ RESULT 19
/ US-08-982-785A-11
/ Sequence 11, Application US/08982785A
/ Patent No. 6258929
/ GENERAL INFORMATION:
/ APPLICANT: Kosik, Kenneth S.
/ APPLICANT: Zhou, Jianhua
/ TITLE OF INVENTION: ALARM RELATED PEPTIDES AND
/ TITLE OF INVENTION: NUCLEIC ACIDS AND DIAGNOSIS USING THEM
/ NUMBER OF SEQUENCES: 13
/ CORRESPONDENCE ADDRESS:
/ ADDRESSEE: Fish & Richardson P.C.
/ STREET: 225 Franklin Street
/ CITY: Boston
/ STATE: MA
/ COUNTRY: USA
/ ZIP: 02110-2804
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Diskette
/ COMPUTER: IBM Compatible
/ OPERATING SYSTEM: Windows 95
/ SOFTWARE: FastSeq for Windows Version 2.0b
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/982,785A
/ FILING DATE: 02-DEC-1997
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US 60/031,556
/ FILING DATE: 02-DEC-1996
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Freeman, John W.
/ REGISTRATION NUMBER: 29,066
/ REFERENCE/DOCKET NUMBER: 05311/018001
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 617/542-5070
/ TELEFAX: 617/542-8906
/ TELEX: 200154
```

INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 666 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-982-785A-11

Query Match 71.4%; Score 30; DB 4; Length 666;
Best Local Similarity 75.0%; Pred. No. 1.3e+02;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
|:||||:|
DB 89 QKIKASRS 96

RESULT 20

US-08-194-211A-3
Sequence 3, Application US/08194211A
Patent No. 5684144

GENERAL INFORMATION:

APPLICANT: ROMEO, TONY

TITLE OF INVENTION: THE ESCHERICHIA COLI ceta GENE, PROTEIN
ENCODED THEREBY AND METHODS OF USE THEREOF

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 Page Mill Road

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/194,211A

FILING DATE: 07-FEB-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: LEHNHARDT, SUSAN K.

REGISTRATION NUMBER: 33,943

REFERENCE/DOCKET NUMBER: 24180-20002.20

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 813-5600

TELEFAX: (415) 494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 61 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-194-211A-3

Query Match 66.7%; Score 28; DB 1; Length 61;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
|:||||:|
DB 49 QRIQAKS 56

RESULT 21

US-08-456-748B-3
Sequence 3, Application US/08456748B
Patent No. 6268471

GENERAL INFORMATION:

APPLICANT: ROMEO, TONY

TITLE OF INVENTION: THE ESCHERICHIA COLI csa GENE, PROTEIN
ENCODED THEREBY AND METHODS OF USE THEREOF

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: MORRISON & FOERSTER

STREET: 755 Page Mill Road

CITY: Palo Alto

STATE: California

COUNTRY: USA

ZIP: 94304-1018

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/456,748B

FILING DATE: 01-JUN-1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: LEHNHARDT, SUSAN K.

REGISTRATION NUMBER: 33,943

REFERENCE/DOCKET NUMBER: 24180-20002.10

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 813-5600

TELEFAX: (415) 494-0792

TELEX: 706141

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 61 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-456-748B-3

Query Match 66.7%; Score 28; DB 4; Length 61;
Best Local Similarity 75.0%; Pred. No. 32;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
|:||||:|
DB 49 QRIQAKS 56

RESULT 22

US-09-442-100-14
Sequence 14, Application US/09442100
Patent No. 6359193

GENERAL INFORMATION:

APPLICANT: Xu, Tian

APPLICANT: Tao, Wufan

APPLICANT: Wang, Weiyl

APPLICANT: Zhang, Sheng

APPLICANT: Yu, Wan

TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCES OF LATS

ENCODED THEREBY AND METHODS BASED THEREON

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/442,100

FILING DATE:

CLASSIFICATION:

```
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/411,111
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Mirock, S. Leslie
; REGISTRATION NUMBER: 18,872
; REFERENCE/DOCKET NUMBER: 6523-003
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 500 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; US-09-442-100-14

Query Match 66.7%; Score 28; DB 4; Length 500;
Best Local Similarity 44.4%; Pred. No. 2.6e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 2 EKVKAARKE 10

RESULT 23
US-08-680-327-3
; Sequence 3, Application US/08680327
; Patent No. 5859321
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, Brian S., Oldroyd, Giles Edward,
; APPLICANT: Sameton, John M., Rommens, Caius
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PLANT
; TITLE OF INVENTION: PATHOGEN RESISTANCE
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klarquist Sparkman Campbell Leigh &
; ADDRESSEE: Whinston
; STREET: One World Trade Center
; STREET: 121 S.W. Salmon Street
; STREET: Suite 1600
; CITY: Portland
; STATE: Oregon
; COUNTRY: United States of America
; ZIP: 97204
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk, 3-1/2 inch
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/680,327
; FILING DATE: July 11, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/310,912
; FILING DATE: September 22, 1994
; CLASSIFICATION: 800
; APPLICATION NUMBER: 08/227,360
; FILING DATE: April 13, 1994
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Dow, Alan. E.
; REGISTRATION NUMBER: 35,123
; REFERENCE/DOCKET NUMBER: 5151-45038
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (503) 226-7391
; TELEFAX: (503) 228-9446
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
```

```
;
; LENGTH: 1824 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-680-327-3

Query Match 66.7%; Score 28; DB 2; Length 1824;
Best Local Similarity 85.7%; Pred. No. 9.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASK 7
Db 119 KRIKASK 125

RESULT 24
US-09-228-246-2
; Sequence 2, Application US/09228246
; Patent No. 6245510
; GENERAL INFORMATION:
; APPLICANT: Staskawicz, B. S. et al.
; TITLE OF INVENTION: PRF Protein and Nucleic Acid Sequences: Compositions
; TITLE OF INVENTION: and Methods for Plant Pathogen Resistance
; FILE REFERENCE: 51700
; CURRENT APPLICATION NUMBER: US/09/228,246
; CURRENT FILING DATE: 1999-01-11
; EARLIER APPLICATION NUMBER: 08/680,327
; EARLIER FILING DATE: 1996-07-11
; EARLIER APPLICATION NUMBER: 08/310,912
; EARLIER FILING DATE: 1994-09-22
; EARLIER APPLICATION NUMBER: 08/227,360
; EARLIER FILING DATE: 1994-04-13
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1824
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
; US-09-228-246-2

Query Match 66.7%; Score 28; DB 4; Length 1824;
Best Local Similarity 85.7%; Pred. No. 9.5e+02;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASK 7
Db 119 KRIKASK 125

RESULT 25
US-09-502-600-106
; Sequence 106, Application US/09502600A
; Patent No. 6294344
; GENERAL INFORMATION:
; APPLICANT: O'Brien, Timothy J.
; TITLE OF INVENTION: Compositions and Methods for the Early Diagnosis of
; TITLE OF INVENTION: Ovarian Cancer
; FILE REFERENCE: D6223CIP-C
; CURRENT FILING DATE: 2000-02-11
; CURRENT APPLICATION NUMBER: US/09/502,600A
; PRIOR APPLICATION NUMBER: 09/039,211
; PRIOR FILING DATE: 03-14-1998
; NUMBER OF SEQ ID NOS: 136
; SEQ ID NO 106
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Residues 88-96 of the SCCE protein
; US-09-502-600-106

Query Match 64.3%; Score 27; DB 4; Length 9;
Best Local Similarity 100.0%; Pred. No. 1.7e+05;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 QRIKAS 6
|||||
Db 4 QRIKAS 9

RESULT 26

US-09-173-581-8
; Sequence 8, Application US/09173581A
; Patent No. 6013455
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina
; APPLICANT: Azimzal, Valda
; APPLICANT: Lu, Aina
; TITLE OF INVENTION: Protein Kinase Homologs
; FILE REFERENCE: PF-0614 US
; CURRENT APPLICATION NUMBER: US/09/173,581A
; CURRENT FILING DATE: 1998-10-15
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 2605059
US-09-173-581-8

Query Match 64.3%; Score 27; DB 3; Length 99;
Best Local Similarity 75.0%; Pred. No. 83;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
|||||
Db 28 QRIAASKA 35

RESULT 27

US-09-420-915-8
; Sequence 8, Application US/09420915
; Patent No. 6264947
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga
; APPLICANT: Tang, Y. Tom
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Guegler, Karl J.
; APPLICANT: Corley, Neil C.
; APPLICANT: Gorgone, Gina
; APPLICANT: Azimzal, Valda
; APPLICANT: Lu, Aina
; TITLE OF INVENTION: Protein Kinase Homologs
; FILE REFERENCE: PF-0614 US
; CURRENT APPLICATION NUMBER: US/09/420,915
; CURRENT FILING DATE: 1999-10-20
; EARLIER FILING DATE: 1998-10-15
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PERL Program
; SEQ ID NO 8
; LENGTH: 99
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE: -
; OTHER INFORMATION: 2605059
US-09-420-915-8

Query Match 64.3%; Score 27; DB 4; Length 99;
Best Local Similarity 75.0%; Pred. No. 83;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
|||||
Db 28 QRIAASKA 35

RESULT 28

US-09-034-916-11
; Sequence 11, Application US/09034916
; Patent No. 6046314
; GENERAL INFORMATION:
; APPLICANT: GEBE, JOHN A.
; APPLICANT: SIADAK, ANTHONY W.
; APPLICANT: ARUFFO, ALEANDRO A.
; TITLE OF INVENTION: SPALPHA: A NOVEL SCAVENGER RECEPTOR
; TITLE OF INVENTION: CYSTEINE-RICH DOMAIN-CONTAINING POLYPEPTIDE AND MONOCLONAL
; TITLE OF INVENTION: ANTIBODIES THERETO
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ROBINS & ASSOCIATES
; STREET: 90 MIDDLEFIELD ROAD, SUITE 200
; CITY: MENLO PARK
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/034,916
; FILING DATE: 04-MAR-1998
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/039,956
; FILING DATE: 06-MAR-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: ROBINS, ROBERTA L.
; REGISTRATION NUMBER: 33,208
; REFERENCE/DOCKET NUMBER: 5998-0020
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 325-7812
; TELEFAX: (650) 325-7823
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 100 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-034-916-11

Query Match 64.3%; Score 27; DB 3; Length 100;
Best Local Similarity 62.5%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 2 RIKASKSF 9
|: ||| |
Db 52 RVNASKGF 59

RESULT 29

5175147-8
; Patent No. 5175147
; APPLICANT: FOLKMAN, MOSES J.; KATO, KOICHI
; TITLE OF INVENTION: ACID-RESISTANT PGF COMPOSITION AND METHOD
; OF TREATING ULCERATING DISEASES OF THE GASTROINTESTINAL TRACT
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/07/382,263
; FILING DATE: 20-JUL-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 234,966
; FILING DATE: 19-AUG-1988
; SEQ ID NO:8:
; LENGTH: 147
5175147-8

Query Match 64.3%; Score 27; DB 6; Length 147;
Best Local Similarity 75.0%; Pred. No. 1.2e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
||:||||
Db 81 QRLASKS 88

RESULT 30
5189147-6
; Patent No. 5189147
; APPLICANT: SAITO, HARUO; KRANZ, DAVID M.; ELSEN, HERMAN N.;
; TONEGAWA, SUSUMU
; TITLE OF INVENTION: METEODIMERIC T LYMPHOCYTE RECEPTOR
; ANTIBODY
; NUMBER OF SEQUENCES: 21
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/271,216
; FILING DATE: 14-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 666,988
; FILING DATE: 31-OCT-1984
; APPLICATION NUMBER: 620,122
; FILING DATE: 13-JUN-1984
; SEQ ID NO:6:
; LENGTH: 202
5189147-6

Query Match 64.3%; Score 27; DB 6; Length 202;
Best Local Similarity 55.6%; Pred. No. 1.7e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
:|:||||
Db 63 KKIEASKDF 71

RESULT 31
US-09-018-576-3
; Sequence 3, Application US/09018576
; Patent No. 5968800
; GENERAL INFORMATION:
; APPLICANT: Gerhold, David L.
; TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, RY60-30
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,576
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hand, J. Mark

; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19885Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732/594-3905
; TELEFAX: 732/594-4720
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-018-576-3

Query Match 64.3%; Score 27; DB 2; Length 325;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
||:||||
Db 254 QRTAASKA 261

RESULT 32
US-09-018-576-12
; Sequence 12, Application US/09018576
; Patent No. 5968800
; GENERAL INFORMATION:
; APPLICANT: Gerhold, David L.
; TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, RY60-30
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/018,576
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hand, J. Mark
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19885Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732/594-3905
; TELEFAX: 732/594-4720
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-018-576-12

Query Match 64.3%; Score 27; DB 2; Length 325;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
||:||||
Db 254 QRTAASKA 261

RESULT 33

US-09-248-137-3
; Sequence 3, Application US/09248137
; Patent No. 6030788
; GENERAL INFORMATION:
; APPLICANT: Gerhold, David L.
; TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, RY60-30
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/248,137
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/018,576
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hand, J. Mark
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19885Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732/594-3905
; TELEFAX: 732/594-4720
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-248-137-3

Query Match 64.3%; Score 27; DB 3; Length 325;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKS 8
Db 254 QRIAASKA 261

RESULT 34
US-09-248-137-12
; Sequence 12, Application US/09248137
; Patent No. 6030788
; GENERAL INFORMATION:
; APPLICANT: Gerhold, David L.
; TITLE OF INVENTION: CYCLIN-DEPENDENT PROTEIN KINASE
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merck & Co., Inc.
; STREET: P.O. Box 2000, RY60-30
; CITY: Rahway
; STATE: NJ
; COUNTRY: US
; ZIP: 07065-0907
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/248,137

; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 09/018,576
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Hand, J. Mark
; REGISTRATION NUMBER: 36,545
; REFERENCE/DOCKET NUMBER: 19885Y
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 732/594-3905
; TELEFAX: 732/594-4720
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 325 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-248-137-12

Query Match 64.3%; Score 27; DB 3; Length 325;
Best Local Similarity 75.0%; Pred. No. 2.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKS 8
Db 254 QRIAASKA 261

RESULT 35
US-09-000-016-7
; Sequence 7, Application US/090000016
; Patent No. 6143541
; GENERAL INFORMATION:
; APPLICANT: Akira ARISAWA et al.
; TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
; TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
; STREET: 2033 K Street, N.W., #800
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Wordperfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/000,016
; FILING DATE: January 30, 1998
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Warren M. Cheek, Jr.
; REGISTRATION NUMBER: 33,367
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-721-8200
; TELEFAX: 202-721-8250
; TELEX:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 520 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear

MOLECULE TYPE: protein
US-09-000-016-7

Query Match 64.3%; Score 27; DB 4; Length 520;
Best Local Similarity 62.5%; Pred. No. 4.4e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKAKSF 9
I: |||:|
Db 42 RVTASKNF 49

RESULT 36

US-09-000-016-4
Sequence 4, Application US/090000016
Patent No. 6143541

GENERAL INFORMATION:

APPLICANT: Akira ARISAWA et al.

TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC

TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DERIV

TITLE OF INVENTION: ITS EXPRESSION PRODUCT

NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

STREET: 2033 K Street, N.W., #800

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/000,016

FILING DATE: January 30, 1998

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200

TELEFAX: 202-721-8250

TELEX:

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 734 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-000-016-4

Query Match 64.3%; Score 27; DB 4; Length 734;
Best Local Similarity 62.5%; Pred. No. 6.2e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKAKSF 9
I: |||:|
Db 246 RVTASKNF 253

RESULT 37

US-09-000-016-2

Sequence 2, Application US/090000016

Patent No. 6143541

GENERAL INFORMATION:

APPLICANT: Akira ARISAWA et al.

TITLE OF INVENTION: GENE ENCODING A PROTEIN HAVING ASYMMETRIC
TITLE OF INVENTION: HYDROLASE ACTIVITY FOR 4-SUBSTITUTED 1,4-DIHYDROPYRIDINE DE
TITLE OF INVENTION: ITS EXPRESSION PRODUCT
NUMBER OF SEQUENCES: 7

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.

STREET: 2033 K Street, N.W., #800

CITY: Washington

STATE: D.C.

COUNTRY: U.S.A.

ZIP: 20006

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/000,016

FILING DATE: January 30, 1998

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

FILING DATE:

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-721-8200

TELEFAX: 202-721-8250

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 823 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

US-09-000-016-2

Query Match 64.3%; Score 27; DB 4; Length 823;

Best Local Similarity 62.5%; Pred. No. 6.9e+02;

Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKAKSF 9
I: |||:|
Db 246 RVTASKNF 253

RESULT 38

US-08-691-641-11

Sequence 11, Application US/08691641

Patent No. 5728819

GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth

APPLICANT: McCoy, John

APPLICANT: Lavallie, Edward

APPLICANT: Racie, Lisa

APPLICANT: Merberg, David

APPLICANT: Treacy, Maurice

APPLICANT: Evans, Cheryl

APPLICANT: Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES

TITLE OF INVENTION: ENCODING THEM

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/691,641
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BROWN, Scott A.
; REGISTRATION NUMBER: 32,724
; TELEPHONE: (617) 498-8224
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-691-641-11

Query Match 61.9%; Score 26; DB 1; Length 65;
Best Local Similarity 71.4%; Pred. No. 89;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 3 IKASKSF 9
Db 57 IKANKTF 63

RESULT 39
US-08-606-789-8
; Sequence 8, Application US/08606789
; Patent No. 5783418
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Stuart, Susan G.
; APPLICANT: Murry, Lynn E.
; APPLICANT: Guegler, Kark J.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/606,789
; FILING DATE: Filed Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/440,743
; FILING DATE: May 5, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/320,011
; FILING DATE: October 5, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0055 US
; TELEPHONE: 415-855-0555

; TELEFAX: 415-852-0195
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; IMMEDIATE SOURCE:
; LIBRARY: GenBank
; CLONE: 163118
; US-08-606-789-8

Query Match 61.9%; Score 26; DB 1; Length 68;
Best Local Similarity 62.5%; Pred. No. 93;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ORIKASKS 8
Db 24 ERIKVSKA 31

RESULT 40
US-09-111-348-8
; Sequence 8, Application US/09111348
; Patent No. 5912130
; GENERAL INFORMATION:
; APPLICANT: Au-Young, Janice
; APPLICANT: Stuart, Susan G.
; APPLICANT: Murry, Lynn E.
; APPLICANT: Guegler, Kark J.
; APPLICANT: Seilhamer, Jeffrey J.
; TITLE OF INVENTION: HUMAN HOMOLOG OF THE RAT G PROTEIN
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: U.S.
; ZIP: 94303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/111,348
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/606,789
; FILING DATE:
; APPLICATION NUMBER: 08/440,743
; FILING DATE: May 5, 1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/320,011
; FILING DATE: October 5, 1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Luther, Barbara J.
; REGISTRATION NUMBER: 33,954
; REFERENCE/DOCKET NUMBER: PF-0055 US
; TELEPHONE: 415-855-0555
; TELEFAX: 415-852-0195
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 68 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

TOPOLGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 163118
US-09-111-348-8

Query Match 61.9%; Score 26; DB 2; Length 68;
Best Local Similarity 62.5%; Pred. No. 93;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
:|||||:
Db 24 ERIKVSKA 31

RESULT 41
PCT-US95-06406A-6
Sequence 6, Application PC/TUS9506406A
GENERAL INFORMATION:
APPLICANT: Janet D. Robshaw, Charles Kunsch
TITLE OF INVENTION: CDNA Clones Encoding Human G Protein
TITLE OF INVENTION: Subunits
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE:
STREET:
CITY:
STATE:
COUNTRY:
ZIP:

COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
COMPUTER: IBM 486
OPERATING SYSTEM: WINDOWS FOR WORKGROUPS
SOFTWARE: WORDPERFECT 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/06406A
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELEPHONE:
TELEFAX:

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 69
TYPE: AMINO ACID
TOPOLOGY: LINEAR
PCT-US95-06406A-6

Query Match 61.9%; Score 26; DB 5; Length 69;
Best Local Similarity 62.5%; Pred. No. 94;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
:|||||:
Db 24 ERIKVSKA 31

RESULT 42
US-08-917-456-2
Sequence 2, Application US/08917456
Patent No. 5868368
GENERAL INFORMATION:
APPLICANT: Greenwood, Rebecca C.
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELEPHONE:
TELEFAX:

TITLE OF INVENTION: NOVEL greA
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dechert Price & Rhoads
STREET: 4000 Bell Atlantic tower, 1717 Arch Stre
CITY: Philadelphia
STATE: PA
COUNTRY: US
ZIP: 19103

COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/917,456
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10046
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:

INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-917-456-2

Query Match 61.9%; Score 26; DB 2; Length 158;
Best Local Similarity 44.4%; Pred. No. 2.2e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
:|||||:
Db 34 ERIKVSF 42

RESULT 43
US-09-229-804-2
Sequence 2, Application US/09229804
Patent No. 6210674
GENERAL INFORMATION:
APPLICANT: Greenwood, Rebecca C.
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELEPHONE:
TELEFAX:

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 69
TYPE: AMINO ACID
TOPOLOGY: LINEAR
PCT-US95-06406A-6

Query Match 61.9%; Score 26; DB 5; Length 69;
Best Local Similarity 62.5%; Pred. No. 94;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKS 8
:|||||:
Db 24 ERIKVSKA 31

RESULT 42
US-08-917-456-2
Sequence 2, Application US/08917456
Patent No. 5868368
GENERAL INFORMATION:
APPLICANT: Greenwood, Rebecca C.
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELEPHONE:
TELEFAX:

FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dickinson, Todd Q
REGISTRATION NUMBER: 28,354
REFERENCE/DOCKET NUMBER: GM10046
TELECOMMUNICATION INFORMATION:
TELEPHONE: 215-994-2252
TELEFAX: 215-994-2222
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 158 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-229-804-2

Query Match 61.9%; Score 26; DB 4; Length 158;
Best Local Similarity 44.4%; Pred. No. 2.2e+02;
Matches 4; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 1 ORIKASKSF 9
Db 34 EKIKVRSF 42

RESULT 44
US-09-032-372-13
Sequence 13, Application US/09032372
Patent No. 6008337
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hillman, Jennifer L.
APPLICANT: Corley, Neil C.
APPLICANT: Guegler, Karl J.
APPLICANT: Yue, Henry
APPLICANT: Lal, Preeti
TITLE OF INVENTION: CELL CYCLE RELATED PROTEINS
NUMBER OF SEQUENCES: 13
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/032,372
FILING DATE: Herewith
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0478 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-845-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:

LIBRARY: GenBank
CLONE: 7345
US-09-032-372-13

Query Match 61.9%; Score 26; DB 3; Length 194;
Best Local Similarity 55.6%; Pred. No. 2.6e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ORIKASKSF 9
Db 149 ERINVSSEF 157

RESULT 45
US-09-078-317-12
Sequence 12, Application US/09078317
Patent No. 6017710
GENERAL INFORMATION:
APPLICANT: Allen, Maxine J.
APPLICANT: Rutter, Marc
APPLICANT: Buckler, Alan J.
TITLE OF INVENTION: RAG Genes and Their Uses
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bozicevic & Reed, LLP
STREET: 285 Hamilton Ave, Suite 200
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94301
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/078,317
FILING DATE: 13-MAY-1998
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Francis, Carol L
REGISTRATION NUMBER: 36,513
REFERENCE/DOCKET NUMBER: SEQ-18P
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-327-3400
TELEFAX: 650-327-3231
TELEX:
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 194 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 6017710e
US-09-078-317-12

Query Match 61.9%; Score 26; DB 3; Length 194;
Best Local Similarity 55.6%; Pred. No. 2.6e+02;
Matches 5; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 ORIKASKSF 9
Db 149 ERINVSSEF 157

Search completed: November 6, 2002, 12:06:34
Job time : 10.1111 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 6, 2002, 12:01:15 ; Search time 23.7778 Seconds
(without alignments)
42.042 Million cell updates/sec

Title: US-09-905-083-99
Perfect score: 42
Sequence: 1 QRIKASKSF 9

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*

```
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	42	100.0	9	AAE08304	Human stratum corn
2	42	100.0	9	AAE08334	Human stratum corn
3	42	100.0	225	AAAB98502	Human stratum corn
4	42	100.0	253	AAAB67888	Human stratum corn
5	42	100.0	253	AAAB05383	Human amyloid prec
6	42	100.0	257	AAAB21326	Human HSCCE. Homo
7	37	88.1	136	ABG23378	Novel human diagno
8	36	85.7	9	AAE08302	Human stratum corn
9	32	76.2	133	AAAG51241	Arabidopsis thalia
10	31	73.8	17	AAAY99042	HLA class II bindi
11	31	73.8	18	AAAY99043	HLA class II bindi

12	31	73.8	18	21	AAAY99044	HLA class II bindi
13	31	73.8	35	19	AAW75017	Human secreted pro
14	31	73.8	111	22	AAAB8837	Human immune/haema
15	31	73.8	114	22	ABBI12260	Human secreted pro
16	31	73.8	114	22	AAAB00066	Human protein SEQ
17	31	73.8	132	21	AAAG01051	Human secreted pro
18	31	73.8	214	22	AAAB79082	Human protein SEQ
19	31	73.8	250	14	AAAR41893	Pcp C. Flavobacte
20	31	73.8	250	17	AAAR99483	Tetrachloro-p-hydr
21	31	73.8	298	22	AAE04777	Human vesicle traf
22	31	73.8	2120	22	ABG20760	Novel human diagno
23	31	73.8	2861	18	AAAB27227	Human TRIO phospho
24	31	73.8	2861	22	ABG17024	Novel human diagno
25	31	73.8	2861	22	ABG68192	Grpase protein SEQ
26	31	73.8	2883	22	ABG06312	Novel human diagno
27	31	73.8	2980	22	ABG20756	Novel human diagno
28	31	73.8	3038	22	ABG17025	Novel human diagno
29	31	73.8	3060	22	ABG06311	Novel human diagno
30	30	71.4	18	21	AAAY99046	HLA class II bindi
31	30	71.4	71	22	AAU20688	Human novel foetal
32	30	71.4	343	21	AAAG20817	Arabidopsis thalia
33	30	71.4	351	21	AAAG39402	Arabidopsis thalia
34	30	71.4	427	21	AAAG20816	Arabidopsis thalia
35	30	71.4	448	21	AAAG20815	Arabidopsis thalia
36	30	71.4	605	21	AAAG39401	Arabidopsis thalia
37	30	71.4	726	21	AAAB36463	Human plakophilin
38	29	69.0	46	22	AAAB94095	Human reproductive
39	29	69.0	79	22	AAAB41295	Human polypeptide
40	29	69.0	127	21	AAAB16734	Bacteriophage Dp-1
41	29	69.0	182	22	ABAB67070	Drosophila melanog
42	29	69.0	203	22	ABAB67070	S. epidermidis ope
43	29	69.0	308	22	ABAB62665	Drosophila melanog
44	29	69.0	438	21	AAAY91342	Group B Streptococ
45	29	69.0	669	22	ABAB69682	Drosophila melanog

ALIGNMENTS

RESULT 1

AAE08304

ID AAE08304 standard; peptide: 9 AA.

XX AC AAE08304;

XX DT 01-NOV-2001 (first entry)

XX DE Human stratum corneum chymotrypsin enzyme peptide #69 (residues 91-99).

XX DE Stratum corneum chymotrypsin enzyme; SCCE: cytostatic; vaccine; tumour;

XX DE cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;

XX DE antitense therapy; malignant hyperplasia.

XX OS Homo sapiens.

XX PN WO200159158-A1.

XX PD 16-AUG-2001.

XX PF 07-FEB-2001; 2001WO-US039977.

XX PR 11-FEB-2000; 2000US-0502600.

XX PR (UYAR-) UNIV ARKANSAS.

XX PI O'Brien TJ;

XX DR WPI; 2001-514676/56.

XX XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin

XX PT enzyme -

XX XX HLA class II bindi

XX PS Claim 25; Page 117; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.
XX

XX Sequence 9 AA;

Query Match 100.0%; Score 42; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 1 QRIKASKSF 9

RESULT 2

AAE08334
ID AAE08334 standard; peptide; 9 AA.

XX AC AAE08334;

XX 01-NOV-2001 (first entry)

XX Human stratum corneum chymotrypsin enzyme peptide #99 (residues 91-99).
DE Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisense therapy; malignant hyperplasia.

XX Homo sapiens.

XX WO200159158-A1.

XX 16-AUG-2001.

XX 07-FEB-2001; 2001WO-US03977.

XX 11-FEB-2000; 2000US-0502600.

XX (UYAR-) UNIV ARKANSAS.

XX O'Brien TJ;

XX WPI; 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin
XX enzyme -

XX Disclosure; Page 124; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by
CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are
CC considered to be an integral part of tumour growth and metastasis, and
CC therefore, markers indicative of their presence or absence are useful
CC for the diagnosis of cancer. The method is useful for diagnosing cancer,
CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used
CC to treat a cancer selected from ovarian, breast, lung, colon, prostate
CC and other cancers in which SCCE is overexpressed. The present sequence is
CC human SCCE peptide.
XX

XX Sequence 9 AA;

Query Match 100.0%; Score 42; DB 22; Length 9;
Best Local Similarity 100.0%; Pred. No. 6.4e+05;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 1 QRIKASKSF 9

RESULT 3

AA98502
ID AAB98502 standard; Protein; 225 AA.

XX AC AAB98502;

XX 03-AUG-2001 (first entry)

XX Human Stratum Corneum Chymotryptic Enzyme, SCCE, catalytic domain.

XX Human; TADG-15; cytostatic; vaccine; ovarian tumour; cancer;
KW tumour antigen-derived gene 15; serine protease;
KW Stratum Corneum Chymotryptic Enzyme; SCCE.

XX Homo sapiens.

XX WO200129056-A1.

XX 26-APR-2001.

XX 20-OCT-2000; 2000WO-US29095.

XX 20-OCT-1999; 99US-0421213.

XX (UYAR-) UNIV ARKANSAS.

XX O'Brien TJ, Tanimoto H;

XX WPI; 2001-381031/40.

XX Novel extracellular serine protease, termed tumor antigen-derived gene
PT 15 protein overexpressed in carcinomas and DNA encoding it, for
PT diagnosis, treatment, prevention of cancer, particularly breast,
PT ovarian cancer -

XX Example 10; Fig 1; 130pp; English.

XX The present invention relates to human tumour antigen-derived gene 15
CC (TADG-15) protein and coding sequence (see AAH23601 and AAB98500).
CC TADG-15 is an extracellular serine protease. It was found that TADG-15 is
CC over-expressed in ovarian tumours. TADG-15 protein or its fragments of
CC 9-20 residues that lack TADG-15 protease activity are useful for
CC vaccinating an individual against TADG-15, having, suspected of having or
CC at risk of getting cancer. Furthermore, the TADG-15 gene can be used as a
CC diagnostic or therapeutic target in cancer. The present sequence was used
CC in a sequence homology alignment with the catalytic domain of TADG-15.

XX Sequence 225 AA;

Query Match 100.0%; Score 42; DB 22; Length 225;
Best Local Similarity 100.0%; Pred. No. 0.22;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 63 QRIKASKSF 71

RESULT 4

AAR67888
ID AAR67888 standard; Protein; 253 AA.

XX AC AAR67888;

XX 09-AUG-1995 (first entry)

XX Human stratum corneum chymotrophic recombinant enzyme (SCCE).
XX

KW Stratum corneum chymotryptic enzyme; skin disorder; acne; psoriasis;
KW callosities; keratosis pilaris; ichthyoses; eczema.
XX Homo sapiens.
XX WO9500651-A.
XX PD 05-JAN-1995.
XX 20-JUN-1994; 94WO-IB00166.
XX PR 18-JUN-1993; 93DK-0000725.
XX (SYMB-) SYMBICOM AB.
XX Egelrud T, Hansson L;
XX WPI: 1995-052088/07.
DR N-PSDB; AAQ81203.
XX
PT Nucleotide sequences encoding stratum corneum chymotryptic enzyme
PT - and related vectors, transformed cells and polypeptides, and
PT useful for treating skin disorders, e.g. acne or psoriasis, and
PT for identification of specific inhibitors.
XX
PS Disclosure; Page 97; 137pp; English.
XX
CC The enzyme encoded by this sequence is used in pharmaceutical, cosmetic
CC and skin care products, especially to treat and prevent acne,
CC xeroderma, or other hyperkeratotic conditions (e.g. callosities or
CC keratosis pilaris), ichthyoses, psoriasis, eczema, etc. It is
CC produced recombinantly following mammal, insect, plant, or
CC microorganism transformation with plasmid pS507.
XX
SQ Sequence 253 AA;

Query Match 100.0%; Score 42; DB 16; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 91 QRIKASKSF 99

RESULT 5
AAW05383
ID AAW05383 standard; Protein; 253 AA.
XX
AC AAW05383;
XX
DT 31-DEC-1996 (first entry)
XX
DE Human amyloid precursor protein protease.
XX
KW Amyloid precursor protein protease; Alzheimer's disease; diagnosis;
KW therapy.
XX
OS Homo sapiens.
XX
XX WO9631122-A1.
XX
PD 10-OCT-1996.
XX
PF 02-APR-1996; 96WO-US04294.
XX
PR 04-APR-1995; 95US-0416257.
XX
PA (ELIL) LILLY & CO ELI.
XX
PI Dixon EP, Johnstone EM, Little SP;
XX
DR WPI; 1996-464694/46.

DR N-PSDB; AAT39783.
XX
PT New isolated human amyloid precursor protein protease - used to
PT develop prods. for the treatment or diagnosis of associated
PT conditions, esp. Alzheimer's disease
XX
XX Claim 1; Page 44-45; 55pp; English.
XX
CC Human amyloid precursor protein protease (AAW05383) is involved in
CC the processing or clearance of amyloid precursor protein to form
CC beta-amyloid peptide. Its amino acid sequence was deduced from
CC a cDNA clone (AAT39783) obt'd. from a human lung library. Recombinant
CC protease can be produced in transformed or transfected prokaryotic
CC (partic. E. coli) or eukaryotic (partic. AV-120 host cells. It is
CC used to develop products for the design and testing of cpds. useful
CC for treating or preventing conditions associated with beta-amyloid
CC peptide, esp. Alzheimer's disease.
XX
SQ Sequence 253 AA;

Query Match 100.0%; Score 42; DB 17; Length 253;
Best Local Similarity 100.0%; Pred. No. 0.25;
Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 91 QRIKASKSF 99

RESULT 6
AAB21326
ID AAB21326 standard; Protein; 257 AA.
XX
AC AAB21326;
XX
DT 02-FEB-2001 (first entry)
XX
DE Human HSCEE.
XX
KW Human; KLK-L1; KLK-L2; KLK-L3; KLK-L4; KLK-L5; KLK-L6; HSCEE;
KW human stratum corneum chymotryptic enzyme; kallikrein-like protein;
KW serine protease; cytostatic; cancer; prostrate cancer.
XX
OS Homo sapiens.
XX
XX WO200053776-A2.
XX
PD 14-SEP-2000.
XX
PF 09-MAR-2000; 2000WO-CA00258.
XX
PR 11-MAR-1999; 99US-0124260.
PR 01-APR-1999; 99US-0127386.
PR 21-JUL-1999; 99US-0144919.
XX
XX (MOUN) MOUNT SINAI HOSPITAL.
XX
PI Yousef GM, Diamandis EP;
XX
XX WPI; 2000-587440/55.
XX
PT New kallikrein-like (KLK-L) proteins for diagnosing and treating KLK-L
PT protein mediated disorders, especially cancer. -
XX
PS Example 4; Fig 17; 184pp; English.
XX
XX The present sequence is human stratum corneum chymotryptic enzyme
XX (HSCEE), a member of the kallikrein multi-gene family. Kallikreins and
CC kallikrein-like proteins are a subgroup of the serine protease enzyme
CC family. They catalyse the selective cleavage of specific polypeptide
CC precursors to release peptides with potent biological activity. Nucleic
CC acids encoding kallikrein-like proteins KLK-L1, KLK-L2, KLK-L3, KLK-L4,
CC KLK-L5 and KLK-L6 have been isolated. The proteins are useful in the

CC treatment, monitoring and diagnosis of cancers, especially prostate
CC cancer. They can also be used to identify a substance that can associate
CC with or mediate the biological activity of the proteins. Antibodies can
CC be used to treat conditions mediated by the kallikrein-like proteins.

SQ Sequence 257 AA;

Query Match 100.0%; Score 42; DB 21; Length 257;

Best Local Similarity 100.0%; Pred. No. 0.25;

Matches 9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

|||||

Db 96 QRIKASKSF 104

RESULT 7

ABG23378
ID: ABG23378 standard; Protein; 136 AA.

XX AC ABG23378;

XX AC ABG23378;

DT 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #23369.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

OS WO200175067-A2.

XX WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX N-PSDB; AAS87565.

XX New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations

PT responsible for genetic disorders or other traits and to assess

PT biodiversity -

XX Claim 20; SEQ ID No 53737; 103pp; English.

PS The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving

CC (II). (II) is useful for generating antibodies against it, detecting or

CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating

CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. ABG00010-ABG30377 represent novel human

CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed

CC

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 136 AA;

Query Match 88.1%; Score 37; DB 22; Length 136;

Best Local Similarity 88.9%; Pred. No. 1.7;

Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

|||||

Db 98 QRIKPSKSF 106

RESULT 8

AAE08302

ID: AAE08302 standard; peptide; 9 AA.

XX AC AAE08302;

XX AC AAE08302;

DT 01-NOV-2001 (first entry)

DE Human stratum corneum chymotrypsin enzyme peptide #67 (residues 90-98).

XX Stratum corneum chymotrypsin enzyme; SCCE; cytostatic; vaccine; tumour;
KW cancer; ovarian; breast; lung; colon; prostate; carcinoma; human;
KW antisenese therapy; malignant hyperplasia.

XX Homo sapiens.

OS WO200159158-A1.

PN 16-AUG-2001.

XX 07-FEB-2001; 2001WO-US03977.

XX 11-FEB-2000; 2000US-0502600.

XX (UYAR-) UNIV ARKANSAS.

XX O'brien TJ;

XX WPI: 2001-514676/56.

XX Diagnosing cancer comprises detecting stratum corneum chymotrypsin

PT enzyme -

XX Disclosure; Page 117; 127pp; English.

XX The invention relates to diagnosing cancer especially ovarian cancer, by

CC screening for stratum corneum chymotrypsin enzyme (SCCE). Proteases are

CC considered to be an integral part of tumour growth and metastasis, and

CC therefore, markers indicative of their presence or absence are useful

CC for the diagnosis of cancer. The method is useful for diagnosing cancer,

CC carcinoma and malignant hyperplasia. The SCCE oligonucleotide may be used

CC to treat a cancer selected from ovarian, breast, lung, colon, prostate

CC and other cancers in which SCCE is overexpressed. The present sequence is

CC human SCCE peptide.

XX Sequence 9 AA;

QY 1 QRIKASKS 8

|||||

Db 2 QRIKASKS 9

RESULT 9

AAG51241

ID: AAG51241 standard; Protein; 133 AA.

Query Match 85.7%; Score 36; DB 22; Length 9;

Best Local Similarity 100.0%; Pred. No. 6.4e+05;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKS 8

|||||

Db 2 QRIKASKS 9

XX AAG51241; 99US-0139462.
XX AC 18-JUN-1999; 99US-0139463.
XX DT 18-JUN-1999; 99US-0139750.
XX DT 18-JUN-1999; 99US-0139763.
XX DE 18-JUN-1999; 99US-0139817.
XX DE 22-JUN-1999; 99US-0139899.
XX DE 23-JUN-1999; 99US-0140353.
XX DE 23-JUN-1999; 99US-0140354.
XX DE 24-JUN-1999; 99US-0140695.
XX DE 28-JUN-1999; 99US-0140823.
XX DE 29-JUN-1999; 99US-0140991.
XX OS Arabidopsis thaliana. 99US-0141287.
XX PN 30-JUN-1999; 99US-0141842.
XX PN 01-JUL-1999; 99US-0142154.
XX PN 02-JUL-1999; 99US-0142055.
XX PD 06-JUL-1999; 99US-0142390.
XX PF 08-JUL-1999; 99US-0142803.
XX PF 09-JUL-1999; 99US-0142920.
XX PF 12-JUL-1999; 99US-0142977.
XX PF 13-JUL-1999; 99US-0143542.
XX PF 14-JUL-1999; 99US-0143624.
XX PF 15-JUL-1999; 99US-0144005.
XX PF 16-JUL-1999; 99US-0144085.
XX PF 16-JUL-1999; 99US-0144086.
XX PF 19-JUL-1999; 99US-0144325.
XX PF 19-JUL-1999; 99US-0144331.
XX PF 19-JUL-1999; 99US-0144332.
XX PF 19-JUL-1999; 99US-0144333.
XX PF 19-JUL-1999; 99US-0144334.
XX PF 19-JUL-1999; 99US-0144335.
XX PF 20-JUL-1999; 99US-0144352.
XX PF 20-JUL-1999; 99US-0144632.
XX PF 21-JUL-1999; 99US-0144884.
XX PF 21-JUL-1999; 99US-0144814.
XX PF 21-JUL-1999; 99US-0145086.
XX PF 21-JUL-1999; 99US-0145088.
XX PF 22-JUL-1999; 99US-0145085.
XX PF 22-JUL-1999; 99US-0145087.
XX PF 22-JUL-1999; 99US-0145089.
XX PF 22-JUL-1999; 99US-0145192.
XX PF 23-JUL-1999; 99US-0145145.
XX PF 23-JUL-1999; 99US-0145218.
XX PF 26-JUL-1999; 99US-0145224.
XX PF 26-JUL-1999; 99US-0145276.
XX PF 27-JUL-1999; 99US-0145913.
XX PF 27-JUL-1999; 99US-0145918.
XX PF 27-JUL-1999; 99US-0145919.
XX PF 28-JUL-1999; 99US-0145951.
XX PF 02-AUG-1999; 99US-0146386.
XX PF 02-AUG-1999; 99US-0146388.
XX PF 02-AUG-1999; 99US-0146389.
XX PF 03-AUG-1999; 99US-0147038.
XX PF 04-AUG-1999; 99US-0147204.
XX PF 04-AUG-1999; 99US-0147302.
XX PF 05-AUG-1999; 99US-0147192.
XX PF 05-AUG-1999; 99US-0147260.
XX PF 06-AUG-1999; 99US-0147303.
XX PF 06-AUG-1999; 99US-0147416.
XX PF 09-AUG-1999; 99US-0147493.
XX PF 09-AUG-1999; 99US-0147935.
XX PF 10-AUG-1999; 99US-0148171.
XX PF 11-AUG-1999; 99US-0148319.
XX PF 12-AUG-1999; 99US-0148341.
XX PF 13-AUG-1999; 99US-0148565.
XX PF 13-AUG-1999; 99US-0148684.
XX PF 16-AUG-1999; 99US-0149368.
XX PF 17-AUG-1999; 99US-0149175.
XX PF 18-AUG-1999; 99US-0149426.
XX PF 20-AUG-1999; 99US-0149722.
XX PF 20-AUG-1999; 99US-0149723.
XX PF 20-AUG-1999; 99US-0149929.
XX PF 23-AUG-1999; 99US-0149902.
XX PF 23-AUG-1999; 99US-0149930.

XX Arabidopsis thaliana protein fragment SEQ ID NO: 65015.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EP1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 23-MAR-1999; 99US-0126264.
XX 23-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.

[illegible]

DT 07-AUG-2000 (first entry)
 XX HLA class II binding antigen epitope peptide #232.
 DE
 XX Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;
 KW immune response; chronic viral disease; cancer; autoimmune disease;
 KW rheumatoid arthritis; multiple sclerosis; myasthenia gravis; AIDS;
 KW allograft rejection; allergy; Lyme disease; hepatitis; prostate cancer;
 KW glomerulonephritis; food hypersensitivity; malaria.
 XX Unidentified.
 OS
 XX WO9961916-A1.
 PN
 XX 02-DEC-1999.
 PD
 XX 28-MAY-1999; 99WO-US12066.
 PF
 XX 29-MAY-1998; 98US-0087192.
 PR
 XX (EPIM-) EPIMMUNE INC.
 PA
 XX Sette A, Southwood S, Sidney J;
 PI
 XX WPI: 2000-097143/08.
 DR
 XX New compositions containing immunogenic peptide epitopes for various
 PT HLA class II DR molecules useful for inducing helper T cell response -
 PS Claim 1; Page 43; 60pp; English.
 CC The present invention relates to a new pharmaceutical composition
 CC comprising a unit dose form of a peptide, or analogue, comprising an
 CC epitope selected from those represented by peptides AAY98812-Y99339
 CC which are derived from various antigens for various human leucocyte
 CC antigen class DR molecules, representative of the world wide population.
 CC The peptide/analogue binds to an HLA class II molecule at an IC-50 of
 CC less than or equal to 1,000 nM. The pharmaceutical can be used to induce
 CC a helper T cell response. The pharmaceutical focuses the immune response
 CC towards selected determinants and could therefore be used in cases of
 CC chronic viral diseases and cancer. Examples of diseases that can be
 CC treated using the peptide containing pharmaceutical include autoimmune
 CC diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia
 CC gravis), allograft rejection, allergies, Lyme disease, hepatitis,
 CC post-streptococcal endocarditis or glomerulonephritis and food
 CC hypersensitivities. The peptide epitopes can be used to enhance immune
 CC responses against other immunogens administered with the peptides.
 CC Diseases which can be treated using immunogenic mixtures include
 CC prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma,
 CC cervical carcinoma, lymphoma, and condyloma acuminatum. The peptides may
 CC also be used to make monoclonal antibodies useful as potential diagnostic
 CC or therapeutic agents. The peptides may also be useful as diagnostic
 CC reagents, for example, to determine the susceptibility of an individual
 CC to a treatment regimen. Also, the peptides may be used to predict which
 CC individuals will be at substantial risk of developing chronic infection.
 CC The selection of appropriate T and B cell epitopes should allow the
 CC development of epitope based vaccines particularly towards conserved
 CC epitopes of pathogens which are characterized by high sequence
 CC variability such as HIV, HCV and Malaria.
 XX
 SQ Sequence 18 AA;
 Query Match 73.8%; Score 31; DB 21; Length 18;
 Best Local Similarity 66.7%; Pred. No. 4.3;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 QRIKASKSF 9
 Db 5 QKIKANSF 13
 RESULT 12
 AAY99044

ID AAY99044 standard; Peptide; 18 AA.
 XX
 AC AAY99044;
 XX
 DT 07-AUG-2000 (first entry)
 XX HLA class II binding antigen epitope peptide #233.
 XX Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;
 KW immune response; chronic viral disease; cancer; autoimmune disease;
 KW rheumatoid arthritis; multiple sclerosis; myasthenia gravis; AIDS;
 KW allograft rejection; allergy; Lyme disease; hepatitis; prostate cancer;
 KW glomerulonephritis; food hypersensitivity; malaria.
 XX Unidentified.
 OS
 XX WO9961916-A1.
 PN
 XX 02-DEC-1999.
 PD
 XX 28-MAY-1999; 99WO-US12066.
 PF
 XX 29-MAY-1998; 98US-0087192.
 PR
 XX (EPIM-) EPIMMUNE INC.
 PA
 XX Sette A, Southwood S, Sidney J;
 PI
 XX WPI: 2000-097143/08.
 DR
 XX New compositions containing immunogenic peptide epitopes for various
 PT HLA class II DR molecules useful for inducing helper T cell response -
 PS Claim 1; Page 43; 60pp; English.
 CC The present invention relates to a new pharmaceutical composition
 CC comprising a unit dose form of a peptide, or analogue, comprising an
 CC epitope selected from those represented by peptides AAY98812-Y99339
 CC which are derived from various antigens for various human leucocyte
 CC antigen class DR molecules, representative of the world wide population.
 CC The peptide/analogue binds to an HLA class II molecule at an IC-50 of
 CC less than or equal to 1,000 nM. The pharmaceutical can be used to induce
 CC a helper T cell response. The pharmaceutical focuses the immune response
 CC towards selected determinants and could therefore be used in cases of
 CC chronic viral diseases and cancer. Examples of diseases that can be
 CC treated using the peptide containing pharmaceutical include autoimmune
 CC diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia
 CC gravis), allograft rejection, allergies, Lyme disease, hepatitis,
 CC post-streptococcal endocarditis or glomerulonephritis and food
 CC hypersensitivities. The peptide epitopes can be used to enhance immune
 CC responses against other immunogens administered with the peptides.
 CC Diseases which can be treated using immunogenic mixtures include
 CC prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma,
 CC cervical carcinoma, lymphoma, and condyloma acuminatum. The peptides may
 CC also be used to make monoclonal antibodies useful as potential diagnostic
 CC or therapeutic agents. The peptides may also be useful as diagnostic
 CC reagents, for example, to determine the susceptibility of an individual
 CC to a treatment regimen. Also, the peptides may be used to predict which
 CC individuals will be at substantial risk of developing chronic infection.
 CC The selection of appropriate T and B cell epitopes should allow the
 CC development of epitope based vaccines particularly towards conserved
 CC epitopes of pathogens which are characterized by high sequence
 CC variability such as HIV, HCV and Malaria.
 XX
 SQ Sequence 18 AA;
 Query Match 73.8%; Score 31; DB 21; Length 18;
 Best Local Similarity 66.7%; Pred. No. 4.3;
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 1 QRIKASKSF 9
 Db 5 QKIKANSF 13

RESULT 13
 AAW75017
 ID AAW75017 standard; Protein; 35 AA.
 XX
 AC AAW75017;
 XX
 DT 19-JAN-1999 (first entry)
 XX
 DE Human secreted protein encoded by gene 176 clone HFTBR48.
 XX
 KW Human; secreted protein; testis; tumour; foetal brain tissue;
 KW fusion protein; cancer; central nervous system; seizure;
 KW diagnosis; neurodegenerative disease.
 XX
 OS Homo sapiens.
 XX
 PN WO9839448-A2.
 XX
 PD 11-SEP-1998.
 XX
 PF 06-MAR-1998; 98WO-US04493.
 XX
 PR 02-OCT-1997; 97US-0061060.
 PR 07-MAR-1997; 97US-0038621.
 PR 07-MAR-1997; 97US-0040161.
 PR 07-MAR-1997; 97US-0040162.
 PR 07-MAR-1997; 97US-0040163.
 PR 07-MAR-1997; 97US-0040333.
 PR 07-MAR-1997; 97US-0040334.
 PR 07-MAR-1997; 97US-0040336.
 PR 07-MAR-1997; 97US-0040626.
 PR 11-APR-1997; 97US-0043311.
 PR 11-APR-1997; 97US-0043312.
 PR 11-APR-1997; 97US-0043313.
 PR 11-APR-1997; 97US-0043314.
 PR 11-APR-1997; 97US-0043568.
 PR 11-APR-1997; 97US-0043569.
 PR 11-APR-1997; 97US-0043576.
 PR 11-APR-1997; 97US-0043578.
 PR 11-APR-1997; 97US-0043580.
 PR 11-APR-1997; 97US-0043669.
 PR 11-APR-1997; 97US-0043670.
 PR 11-APR-1997; 97US-0043671.
 PR 11-APR-1997; 97US-0043672.
 PR 11-APR-1997; 97US-0043674.
 PR 23-MAY-1997; 97US-0047492.
 PR 23-MAY-1997; 97US-0047500.
 PR 23-MAY-1997; 97US-0047501.
 PR 23-MAY-1997; 97US-0047502.
 PR 23-MAY-1997; 97US-0047503.
 PR 23-MAY-1997; 97US-0047581.
 PR 23-MAY-1997; 97US-0047582.
 PR 23-MAY-1997; 97US-0047583.
 PR 23-MAY-1997; 97US-0047584.
 PR 23-MAY-1997; 97US-0047585.
 PR 23-MAY-1997; 97US-0047586.
 PR 23-MAY-1997; 97US-0047587.
 PR 23-MAY-1997; 97US-0047588.
 PR 23-MAY-1997; 97US-0047589.
 PR 23-MAY-1997; 97US-0047590.
 PR 23-MAY-1997; 97US-0047592.
 PR 23-MAY-1997; 97US-0047593.
 PR 23-MAY-1997; 97US-0047594.
 PR 23-MAY-1997; 97US-0047595.
 PR 23-MAY-1997; 97US-0047596.
 PR 23-MAY-1997; 97US-0047597.
 PR 23-MAY-1997; 97US-0047598.
 PR 23-MAY-1997; 97US-0047599.
 PR 23-MAY-1997; 97US-0047600.
 PR 23-MAY-1997; 97US-0047601.
 PR 23-MAY-1997; 97US-0047612.

PR 23-MAY-1997; 97US-0047613.
 PR 23-MAY-1997; 97US-0047614.
 PR 23-MAY-1997; 97US-0047615.
 PR 23-MAY-1997; 97US-0047617.
 PR 23-MAY-1997; 97US-0047618.
 PR 23-MAY-1997; 97US-0047632.
 PR 23-MAY-1997; 97US-0047633.
 PR 06-JUN-1997; 97US-0048964.
 PR 06-JUN-1997; 97US-0048974.
 PR 13-JUN-1997; 97US-0049610.
 PR 08-JUL-1997; 97US-0051926.
 PR 16-JUL-1997; 97US-0052874.
 PR 18-AUG-1997; 97US-0055724.
 PR 22-AUG-1997; 97US-0056630.
 PR 22-AUG-1997; 97US-0056631.
 PR 22-AUG-1997; 97US-0056632.
 PR 22-AUG-1997; 97US-0056636.
 PR 22-AUG-1997; 97US-0056637.
 PR 22-AUG-1997; 97US-0056662.
 PR 22-AUG-1997; 97US-0056664.
 PR 22-AUG-1997; 97US-0056845.
 PR 22-AUG-1997; 97US-0056862.
 PR 22-AUG-1997; 97US-0056864.
 PR 22-AUG-1997; 97US-0056872.
 PR 22-AUG-1997; 97US-0056874.
 PR 22-AUG-1997; 97US-0056875.
 PR 22-AUG-1997; 97US-0056876.
 PR 22-AUG-1997; 97US-0056877.
 PR 22-AUG-1997; 97US-0056878.
 PR 22-AUG-1997; 97US-0056879.
 PR 22-AUG-1997; 97US-0056880.
 PR 22-AUG-1997; 97US-0056881.
 PR 22-AUG-1997; 97US-0056882.
 PR 22-AUG-1997; 97US-0056884.
 PR 22-AUG-1997; 97US-0056886.
 PR 22-AUG-1997; 97US-0056887.
 PR 22-AUG-1997; 97US-0056888.
 PR 22-AUG-1997; 97US-0056889.
 PR 22-AUG-1997; 97US-0056892.
 PR 22-AUG-1997; 97US-0056893.
 PR 22-AUG-1997; 97US-0056894.
 PR 22-AUG-1997; 97US-0056903.
 PR 22-AUG-1997; 97US-0056908.
 PR 22-AUG-1997; 97US-0056909.
 PR 22-AUG-1997; 97US-0056910.
 PR 05-SEP-1997; 97US-0056911.
 PR 05-SEP-1997; 97US-0057650.
 PR 05-SEP-1997; 97US-0057669.
 PR 05-SEP-1997; 97US-0057761.
 PR 12-SEP-1997; 97US-0058785.
 XX
 (HUMA-) HUMAN GENOME SCI INC.
 PA Bednarik DP, Brewer LA, Carter KC, Duan R, Ebner R, Endress CA;
 PI Feng P, Ferrie AM, Fischer CL, Florence KA, Greene JM, Hu JS;
 PI Kyaw H, Lafleur DM, Li Y, Moore PA, Ni J, Olsen HS, Rosen CA;
 PI Ruben SM, Shi Y, Soppet DR, Young PE, Yu GL, Zeng Z;
 XX
 WPI: 1998-506364/43.
 DR N-PSDB; AAV59803.
 XX
 PT New isolated human genes and the secreted polypeptide(s) they encode
 PT - useful for diagnosis and treatment of e.g. cancers, neurological
 PT disorders, immune diseases, inflammation or blood disorders
 XX
 PS Claim 1; Page 701; 721pp; English.
 XX
 CC This sequence represents a secreted human protein encoded by the nucleic
 CC acid molecule designated Gene 176 from the human cDNA clone HFTBR48
 CC (deposited as clone ATCC 97904 and ATCC 209050).
 CC The gene can be used to generate fusion proteins by linking to the gene
 CC to a human immunoglobulin Fc portion (e.g. AAV59502) for increasing the
 CC stability of the fused protein as compared to the human protein only.

CC The invention relates to 186 novel genes and their fragments (nucleic
CC acid sequences: AAV59511-V59812; amino acid sequences AAW74731-W75036)
CC which are useful for preventing, treating or ameliorating medical
CC conditions e.g. by protein or gene therapy. Also, pathological
CC conditions can be diagnosed by determining the amount of the new
CC polypeptides in a sample or by determining the presence of mutations in
CC the new polynucleotides. Specific uses are described for each of the 186
CC polynucleotides, based on which tissues they are most highly expressed in
CC (see AAV59511 for described uses).

XX Sequence 35 AA;

Query Match 73.8%; Score 31; DB 19; Length 35;

Best Local Similarity 85.7%; Pred. No. 8.7;

Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 IKASKSF 9

:|||||

Db 18 VKASKSF 24

RESULT 14

AAW88837

ID AAW88837 standard; Protein; 111 AA.

XX AC AAW88837;

XX DT 07-NOV-2001 (first entry)

XX DE Human Immune/haematopoietic antigen SEQ ID NO:16430.

XX KW Human; Immune; haematopoietic; immune/haematopoietic antigen; cancer;
XX cytotstatic; gene therapy; vaccine; metastasis.

XX OS Homo sapiens.

XX PN WO200157182-A2.

XX PD 09-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01354.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0216647.

XX PR 11-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 14-JUL-2000; 2000US-0217496.

XX PR 26-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 14-AUG-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225466.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 25-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX
PI Rosen CA, Barash SC, Ruben SM;
DR N-PSDB; AAK61618.
DR WPI; 2001-483426/52.
DR N-PSDB; AAK61618.
XX
XX
PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT useful for preventing, diagnosing and/or treating cancers and
PT metastasis -
XX
XX
XX Claim 11; SEQ ID NO 16430; 3071pp + Sequence Listing; English.
PS
PS AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
CC activity, and can be used in gene therapy and vaccine production. (I)
CC proteins and polynucleotides may be used in the prevention, diagnosis and
CC treatment of diseases associated with inappropriate (I) expression. For
CC example, they may be used to treat disorders associated with decreased
CC expression by rectifying mutations or deletions in a patient's genome
CC that affect the activity of (I) by expressing inactive proteins or to
CC supplement the patient's own production of (I). Additionally, (I)
CC polynucleotides may be used to produce the secreted (I), by inserting
CC the nucleic acids into a host cell and culturing the cell to express the
CC protein. (I) proteins and polynucleotides may be used to prevent,
CC diagnose and treat immune/hematopoietic-related diseases, especially
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
CC to AAK87694 represent human immune/hematopoietic antigen genomic
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
CC represent sequences used in the exemplification of the present invention.
XX
XX
SQ Sequence 111 AA;
Query Match 73.8%; Score 31; DB 22; Length 111;
Best Local Similarity 75.0%; Pred. No. 29;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 2 RIKASKSF 9
Db 98 RLKASESF 105

RESULT 15

ABBI12260

ID ABBI12260 standard; peptide: 114 AA.

XX

AC ABBI12260;

XX

DT 11-JAN-2002 (first entry)

XX

DE Human secreted protein homologue, SEQ ID NO:2630.

XX

KW Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW inhibitor; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW myeloid cell disorder; lymphoid cell disorder; arthritis;
KW chronic inflammatory condition; proliferative retinopathy;
KW atherosclerosis; coronary heart disease; arterial ischaemia;
KW bone disorder; osteoporosis; vascular growth disorder;
KW tissue regeneration; wound healing; infection; immune disorder;
KW cell culture; drug screening; gene therapy; antiinflammatory;
KW antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW antifungal; vulnerary; antiulcer.

XX

OS Homo sapiens.

XX

PN WO200157188-A2.

XX

PD 09-AUG-2001.

XX

PF 05-FEB-2001; 2001WO-US03800.

XX

PR 03-FEB-2000; 2000US-0496914.

PR

PR 27-APR-2000; 2000US-0560875.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

WPI; 2001-457740/49.

DR

N-PSDB; ABA09504.

XX

Human proteins and DNA encoding sequences useful for preventing,

PT

treating or ameliorating a medical condition in a mammalian subject

PT

e.g. arthritis and cancer -

PS

Claim 20; Page 321; 1963pp; English.

XX

Sequences ABBI10981-ABBI12330 represent 1350 novel human polypeptides, and
sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
invention also relates to vectors and recombinant host cells comprising a
nucleotide of the invention, methods of producing the novel polypeptides,
antibodies against the polypeptides, methods of detecting the nucleotides
or polypeptides in a sample, and methods of identifying compounds which
bind to polypeptides of the invention. Although novel, many of the
polypeptides of the invention have homology to known proteins, thereby
giving an insight into their probable biological activities, and hence
potential therapeutic applications. The polypeptides of the invention may
have various activities, including cytokine, cell proliferation or cell
differentiation activities; stem cell growth factor activity;
haematopoiesis regulatory activity; tissue growth activity;
immunomodulatory activity; activin- or inhibin-related activities;
chemotactic or chemokinetic activities; haemostatic, thrombotic or
thrombolytic activities; receptor or ligand activities; or may be
involved in oncogenesis, cancer cell proliferation or metastasis.
Depending on their biological activities, polypeptides and nucleotides of
the invention are useful for preventing, treating or ameliorating medical
conditions, e.g., by protein or gene therapy. Such conditions include
cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
proliferative retinopathy, atherosclerosis, coronary heart disease,

arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal vascular growth. Polypeptides involved with tissue regeneration and repair (or nucleic acids encoding them) may be used to promote wound healing (e.g., of burns, incisions and ulcers), while those with immunomodulatory activities may be used in the treatment of viral, bacterial and fungal infections in addition to immune disorders. Polypeptides with growth factor activity may be used in cell cultures to promote cell growth. For example, such polypeptides may be used to manipulate stem cells in culture to give rise to neuroepithelial cells that can be used to augment or replace cells damaged by illness, autoimmune disease or accidental damage. The polypeptides and nucleotides may also be used in the diagnosis of the above conditions, and in drug screening techniques. The present sequence represents a novel human polypeptide of the invention.

Query Match 73.8%; Score 31; DB 22; Length 114;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9
Db 22 KRVKASHSF 30

RESULT 16
AAM80066
ID AAM80066 standard; Protein; 114 AA.

AC AAM80066;
DT 06-NOV-2001 (first entry)
DE Human protein SEQ ID NO 3712.

Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorder; arthritis; inflammation.

Homo sapiens.
WO200157190-A2.
09-AUG-2001.
05-FEB-2001; 2001WO-US04098.
03-FEB-2000; 2000US-0496914.
27-APR-2000; 2000US-0560875.
20-JUN-2000; 2000US-0598075.
19-JUL-2000; 2000US-0620325.
01-SEP-2000; 2000US-0654936.
15-SEP-2000; 2000US-0663561.
20-OCT-2000; 2000US-0693325.
30-NOV-2000; 2000US-0728422.
(HYSE-) HYSEQ INC.

Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
Xue AJ, Yang Y, Wejhrman T, Goodrich R;
WPI; 2001-476283/51.
N-PSDB; AAK53199.
Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy -
Claim 20; Page 416; 6221pp; English.
The invention relates to polynucleotides (AAK51456-AAK53435) and the

encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or treatment of cancer, leukaemia, nervous system disorders, arthritis and inflammation.

Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the sequence listing were missing at the time of publication.

Query Match 73.8%; Score 31; DB 22; Length 114;
Best Local Similarity 66.7%; Pred. No. 30;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9
Db 22 KRVKASHSF 30

RESULT 17
AAG01051
ID AAG01051 standard; Protein; 132 AA.

AC AAG01051;
DT 06-OCT-2000 (first entry)
DE Human secreted protein, SEQ ID NO: 5132.

Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation; gene therapy; chromosome mapping.

Homo sapiens.
EP1033401-A2.
06-SEP-2000.
21-FEB-2000; 2000EP-0200610.
26-FEB-1999; 99US-0122487.
(GEST) GENSET.
Dumas Milne Edwards J, Duclert A, Giordano J;
WPI; 2000-500381/45.
N-PSDB; AAC01057.
New nucleic acid that is a 5' expressed sequence tag (5' EST) for obtaining cDNAs and genomic DNAs that correspond to 5' ESTs and for diagnostic, forensic, gene therapy and chromosome mapping procedures -
Claim 13; SEQ ID 5132; 71pp + CD-ROM; English.

The present sequence is a polypeptide encoded by one of a large number of 5' ESTs derived from total human RNAs or polyA+ RNAs derived from 30 different tissues. EST sequences usually correspond mainly to the 3' untranslated region (UTR) of the mRNA because they are often obtained from oligo-dT primed cDNA libraries. Such ESTs are not well suited for isolating cDNA sequences derived from the 5' ends of mRNAs and even in those cases where longer cDNA sequences have been obtained, the full 5' UTR is rarely included. 5' ESTs are derived from mRNAs with intact 5' ends and can therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in diagnostic, forensic, gene therapy and chromosome mapping procedures. They are used to obtain upstream

CC regulatory sequences and to design expression and secretion vectors.

```
XX SQ Sequence 132 AA;
Query Match 73.8%; Score 31; DB 21; Length 132;
Best Local Similarity 66.7%; Pred. No. 35;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 19 KRVKASHSF 27
:|:|:|:|

RESULT 18
AAM79082
ID AAM79082 standard; Protein; 214 AA.
XX
AC AAM79082;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 1744.
XX
KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
WPI: 2001-476283/51.
N-PSDB; AAK52215.
XX
CC Nucleic acids encoding polypeptides with cytokine-like activities,
useful in diagnosis and gene therapy -
XX
Claim 20; Page 4079; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
```

```
XX SQ Sequence 214 AA;
Query Match 73.8%; Score 31; DB 22; Length 214;
Best Local Similarity 66.7%; Pred. No. 58;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 19 KRVKASHSF 27
:|:|:|:|

RESULT 19
AAR41893
ID AAR41893 standard; Protein; 250 AA.
XX
AC AAR41893;
XX
DT 20-APR-1994 (first entry)
XX
DE Pcp C.
XX
KW Penta-chlorophenol; PCP; degradation; contamination; fungicide;
KW pesticide; herbicide; disinfectant.
XX
OS Flavobacterium sp. (Strain ATCC 39723).
XX
PN WO9319169-A.
XX
PD 30-SEP-1993.
XX
PF 19-MAR-1993; 93WO-US02460.
XX
PR 23-MAR-1992; 92US-0856015.
PR 13-JUL-1992; 92US-0914282.
XX
PA (IDAH-) IDAHO RES FOUND INC.
XX
PI Orser CS, Xun L;
XX
WPI: 1993-320736/40.
DR N-PSDB; AAK49738.
XX
PT Novel protein PcpA, PcpB, and PcpC from Flavobacterium sp. Strain
ATCC 39723 - involves in breakdown of penta:chlorophenol, for
PT bio-remedial applications
XX
Claim 17; Page 73-74; 102pp; English.
XX
CC Penta-chlorophenol (PCP) is a component of fungicides, pesticides,
CC herbicides and disinfectants. It is highly toxic and persists
CC within the environment and within food chains. PcpA, PcpB and PcpC
CC degrade PCP. PcpB catalyses the conversion of PCP to 2,3,5,6-
CC tetrachloro-p-hydroquinone (TeCH) and PcpC catalyses the conversion
CC of TeCH to 2,3,6-trichloro-hydroquinone (TrCH) and then to 2,6-
CC dichloro-p-hydroquinone. The enzymes can be used to remove
CC contamination PCP from the environment.
XX
SQ Sequence 250 AA;
Query Match 73.8%; Score 31; DB 14; Length 250;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
Db 229 QRMKARRSF 237
||:|:|:|

RESULT 20
AAR99483
ID AAR99483 standard; Protein; 250 AA.
XX
```


AC AAR99483;
XX
DT 25-SEP-1996 (first entry)
XX
DE Tetrachloro-p-hydroquinone reductase.
XX
KW Pentachlorophenol breakdown pathway; PcpC; PcpA; PcpB; environment;
KW Tetrachloro-p-hydroquinone reductase; PCP-degrading enzyme complex;
KW Flavobacterium; PCP; 2,3,5,6-tetrachloro-p-hydroquinone; Tech;
KW glutathione; 2,3,6-trichloro-p-hydroquinone; TrCH; food chain;
KW 2,6-dichloro-p-hydroquinone; wood preserving industry; fungicide;
KW pesticide; herbicide; disinfectant.
XX
OS Flavobacterium sp. Strain ATCC 39723.
XX
PN US5512478-A.
XX
PD 30-APR-1996.
XX
PF 23-MAR-1992; 92US-0856015.
XX
PR 18-JUL-1994; 94US-0276887.
PR 23-MAR-1992; 92US-0856015.
PR 13-JUL-1992; 92US-0914282.
XX
PA (IDAH-) IDAHO RES FOUND INC.
XX
XN Lange CC, Orser CS, Xun L;
XX
DR WPI; 1996-229872/23.
DR N-PSDB; AAT34182.
XX
XX Flavobacterium sp. penta:chloro:phenol breakdown pathway genes and
PT enzymes - useful in the bio-remediation and dechlorination of PCP
PT contg. matter
XX
PS Claim 2; Columns 49-52; 52pp; English.
XX
CC The sequences given in AAR99483-85 represent proteins from the
CC pentachlorophenol (Pcp) breakdown pathway, designated PcpC, PcpA and
CC PcpB respectively. This sequence represents the protein tetrachloro-p-
CC hydroquinone reductase. PcpA has a mol. wt. of approx. 30000 and is
CC thought to be involved in PCP breakdown either at the stage of PCP
CC uptake into the cell or as a component of a PCP-degrading enzyme
CC complex. PcpB and PcpC are enzymes which catalyse successive steps
CC in the PCP breakdown pathway of Flavobacterium. PcpB catalyses the
CC conversion of PCP to 2,3,5,6-tetrachloro-p-hydroquinone (Tech) in
CC the presence of NADPH and oxygen. It has a mol. wt. of 59-63 kD.
CC In the presence of the reduced form of glutathione, PcpC catalyses the
CC conversion of Tech to 2,3,6-trichloro-p-hydroquinone (TrCH) and the
CC conversion of TrCH to 2,6-dichloro-p-hydroquinone. Purified PcpC has
CC a mol. wt. of 26-29 kD. In combination these enzymes catalyse the
CC initial steps of PCP breakdown. These enzymes can specifically be
CC used in the breakdown of PCP containing matter which persists in the
CC environment and in food chains after its use in the wood preserving
CC industry as a fungicide and pesticide, and in products such as
CC herbicides and disinfectants.
XX
SQ Sequence 250 AA;
Query Match 73.8%; Score 31; DB 17; Length 250;
Best Local Similarity 66.7%; Pred. No. 68;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 QRIKASKSF 9
Db 229 QRKARRSF 237
||:||||
RESULT 21
AAE04777
ID AAE04777 standard; Protein; 298 AA.
XX

AC AAE04777;
XX
DT 10-SEP-2001 (first entry)
XX
DE Human vesicle trafficking protein-20 (VETRP-20) protein.
XX
KW Human; vesicle trafficking protein-20; VETRP-20; cystic fibrosis; goitre;
KW glucose-galactose malabsorption syndrome; hypercholesterolaemia; vaccine;
KW diabetes mellitus; diabetes insipidus; hyperglycaemia; hypoglycaemia;
KW Grave's disease; Cushing's disease; Addison's disease; AIDS; allergy;
KW ulcerative colitis; gastrointestinal disorder; asthma; hay fever; gout;
KW autoimmune disease; inflammatory disease; bowel disease; osteoporosis;
KW multiple sclerosis; rheumatoid arthritis; psoriasis; anaemia; cancer;
KW pancreatitis; Crohn's disease; glomerulonephritis; atherosclerosis;
KW Goodpasture's syndrome; Hashimoto's thyroiditis; gene therapy; virucide;
KW systemic lupus erythematosus; dermatitis; nephrotropic; antihelminthic;
KW cerebroprotective.
XX
OS Homo sapiens.
XX
PN WO200146256-A2.
XX
PD 28-JUN-2001.
XX
PF 21-DEC-2000; 2000WO-US34919.
XX
PR 21-DEC-1999; 99US-0172968.
PR 23-DEC-1999; 99US-0172066.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Tang YT, Yue H, Bandman O, Hillman JL, Baughn MR, Lu DAM;
PI Azimzai Y, Yang J, Burford N, Au-Young J, Reddy R;
XX
DR WPI; 2001-418040/44.
DR N-PSDB; AAD09390.
XX
XX Novel human vesicle trafficking proteins useful for treating and
PT preventing vesicle trafficking disorders, autoimmune/inflammatory
PT disorders and cancers -
XX
PS Claim 1; Page 123-124; 144pp; English.
XX
CC The present sequence is human vesicle trafficking protein-20 (VETRP-20)
CC protein. VETRP is used as vaccine. VETRP is useful for treating a disease
CC or condition associated with decreased expression of functional VETRP,
CC such as vesicle trafficking disorders e.g., cystic fibrosis, glucose-
CC galactose malabsorption syndrome, hypercholesterolaemia, diabetes
CC mellitus, diabetes insipidus, hyperglycaemia, hypoglycaemia, Grave's
CC disease, goitre, Cushing's disease, Addison's disease, gastrointestinal
CC disorders including ulcerative colitis, AIDS, allergies including
CC asthma, hay fever, autoimmune/inflammatory diseases including
CC inflammatory bowel disease, multiple sclerosis, rheumatoid arthritis,
CC osteoporosis, viral, bacterial, fungal, helminthic and protozoal
CC infections, psoriasis, pancreatitis, anaemia, Crohn's disease,
CC glomerulonephritis, atherosclerosis, dermatitis, Hashimoto's
CC thyroiditis, gout, Goodpasture's syndrome, systemic lupus erythematosus
CC and cancers. VETRP polynucleotides are useful in gene therapy and in
CC diagnostic purposes.
XX
SQ Sequence 298 AA;
Query Match 73.8%; Score 31; DB 22; Length 298;
Best Local Similarity 66.7%; Pred. No. 82;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
QY 1 QRIKASKSF 9
Db 19 KRKAKSHSF 27
:|:||||
RESULT 22
ABG20760

ID ABG20760 standard; Protein; 2120 AA.
 AC
 AC ABG20760;
 DT
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #20751.
 DE
 DE Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 KW
 KW Homo sapiens.
 OS
 OS WO200175067-A2.
 PN
 PN 11-OCT-2001.
 PD
 PD 30-MAR-2001; 2001WO-US08631.
 XX
 XX 31-MAR-2000; 2000US-0540217.
 PR
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 PA Drmanac RT, Liu C, Tang YT;
 PI
 PI WPI: 2001-639362/73.
 XX
 XX N-PSDB; AAS84947.
 DR
 DR New isolated polynucleotide and encoded polypeptides, useful in
 XX diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PT
 XX Claim 20; SEQ ID No 51119; 103pp; English.
 PS
 PS The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2120 AA;
 Query Match 73.8%; Score 31; DB 22; Length 2120;
 Best Local Similarity 56.7%; Pred. No. 6.4e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QRIKASKSF 9
 Db 173 QRIQSSESF 181
 RESULT 23
 AAW27227
 ID AAW27227 standard; Protein; 2861 AA.
 XX

AC AAW27227;
 XX
 DT 27-MAR-1998 (first entry)
 XX
 DE Human TRIO phosphoprotein.
 XX
 DE TRIO; human; proliferative disease; neurodegenerative disease;
 KW oncogene; cytoskeleton; actin; diagnosis; transgenic animal;
 KW antibody; therapy; ss.
 KW
 OS Homo sapiens.
 OS
 OS Key Location/Qualifiers
 FH 252..359
 FT Domain /label= SP-D1
 FT /note= "putative spectrin repeat sequence"
 FT 479..585
 FT Domain /label= SP-D2
 FT /note= "putative spectrin repeat sequence"
 FT 819..925
 FT Domain /label= SP-D3
 FT /note= "putative spectrin repeat sequence"
 FT 1050..1157
 FT Domain /label= SP-D4
 FT /note= "putative spectrin repeat sequence"
 FT 1237..1407
 FT Domain /label= GEF-D1
 FT /note= "GEF domain 1, has rac-1 specificity"
 FT 1435..1534
 FT Domain /label= PSK
 FT /note= "pleckstrin homology domain"
 FT 1914..2085
 FT Domain /label= GEF-D2
 FT /note= "GEF domain 2, has rhoa specificity"
 FT 2113..2214
 FT Domain /label= PSK
 FT /note= "pleckstrin homology domain"
 FT 2448..2541
 FT Domain /note= "Ig-like domain"
 FT 2560..2816
 FT Domain /note= "serine/threonine kinase domain"
 FT
 PN WO9735979-A1.
 XX
 PD 02-OCT-1997.
 XX
 PF 27-MAR-1997; 97WO-US052336.
 XX
 PR 27-MAR-1996; 96US-0014214.
 XX
 PA (DAND) DANA FARBBER CANCER INST INC.
 XX
 PI Debant A, Serra-Pages C, Streuli M;
 XX WPI: 1997-489642/45.
 DR N-PSDB; AAT85392.
 DR
 PT New nucleic acid encoding the TRIO phosphoprotein - used for
 PT diagnosis and treatment of proliferative and neurodegenerative
 PT diseases
 PS Claim 2; Page 93-103; 141pp; English.
 XX
 CC This protein sequence comprises the novel multidomain human
 CC phosphoprotein TRIO, which is suggested to be a central organiser of
 CC multiple signalling pathways, to be involved in the activation of
 CC oncogenes such as c-fos, and to induce transformation of cells.
 CC The amino acid sequence was deduced from a composite of cDNA clones
 CC obtained from human WI-38 fibroblast cDNA by interaction trap assay.
 CC TRIO mRNA was expressed in all 8 human tissues examined. Once
 CC isolated, TRIO DNA can be expressed in usual host/vector systems to
 CC produce recombinant TRIO polypeptides, e.g. for antibody production
 CC or screening for modulators of activity. Labelled antibodies that

CC bind TRIO specifically can be used to detect/quantify TRIO activity,
 CC particularly for diagnosis and phenotyping of neoplastic or
 CC hyperplastic disease. Modulators of TRIO activity (e.g. antisense
 CC nucleic acids, antibodies, peptides or mimetics) can be used to
 CC reorganise the actin cytoskeleton (claimed), e.g. in cases of wound
 CC healing and/or tumour metastasis, to treat an oncogene (claimed),
 CC or more generally to control growth, differentiation, migration
 CC and/or survival of cells, e.g. regulation of the immune response to
 CC infection, treatment of impaired immune response (as in chronic
 CC granulomatous disease), control of apoptosis in cancer therapy, and
 CC treatment of degenerative diseases (e.g. Parkinson's, Alzheimer's
 CC or Huntington's, amyotrophic lateral sclerosis, gastric ulcers,
 CC Wilms' tumour etc.). Transgenic animals can be used as models to
 CC characterise TRIO genes and proteins.
 XX
 SQ Sequence 2861 AA;
 Query Match 73.8%; Score 31; DB 18; Length 2861;
 Best Local Similarity 66.7%; Pred. No. 8.7e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ORIKASKSF 9
 |||::|:|
 Db 225 QRIQSSESF 233
 RESULT 24
 ABG17024
 ID ABG17024 standard; Protein; 2861 AA.
 XX
 AC ABG17024;
 XX
 DT 18-FEB-2002 (first entry)
 XX
 DE Novel human diagnostic protein #17015.
 XX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 KW Homo sapiens.
 OS
 XX
 XX WO200175067-A2.
 PN
 XX
 PD 11-OCT-2001.
 XX
 XX 30-MAR-2001; 2001WO-US08631.
 PF
 XX 31-MAR-2000; 2000US-0540217.
 PR
 PR 23-AUG-2000; 2000US-0649167.
 XX
 XX (HYSE-) HYSEQ INC.
 PA
 XX
 XX Drmanac RT, Liu C, Tang YT;
 PI
 XX
 XX WPI; 2001-639362/73.
 DR
 DR N-PSDB; AAS81211.
 XX
 XX New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity -
 PT
 XX
 XX Claim 20; SEQ ID No 47383; 103pp; English.
 PS
 XX The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as

CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2861 AA;
 Query Match 73.8%; Score 31; DB 22; Length 2861;
 Best Local Similarity 66.7%; Pred. No. 8.7e+02;
 Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ORIKASKSF 9
 |||::|:|
 Db 225 QRIQSSESF 233
 RESULT 25
 AAG68192
 ID AAG68192 standard; Protein; 2861 AA.
 XX
 AC AAG68192;
 XX
 DT 25-JAN-2002 (first entry)
 XX
 DE GTPase protein SEQ ID NO:108.
 XX
 KW Human; high bone mass; HBM gene; Zmax1 gene; chromosome 11; 11q13.3;
 KW sequence tagged site; STS; osteoporosis; osteopathic; gene therapy;
 KW antisense therapy; vaccine; bone disorder; Paget's disease;
 KW sclerostosis; osteomalacia; fibrous dysplasia.
 XX
 OS Homo sapiens.
 XX
 XX WO200177327-A1.
 PN
 XX
 PD 18-OCT-2001.
 XX
 XX 21-JUN-2000; 2000WO-US16951.
 PF
 XX 05-APR-2000; 2000US-0543771.
 PR
 PR 05-APR-2000; 2000US-0544398.
 XX
 XX (GENO-) GENOME THERAPEUTICS CORP.
 PA
 XX
 XX Carulli JP, Little RD, Recker RR, Johnson ML;
 PI
 XX WPI; 2001-657171/75.
 DR
 XX
 XX New high bone mass (HBM) and Zmax1 genes and proteins useful for
 PT modulating bone mass for the treatment of e.g. osteoporosis -
 PT
 XX
 XX Claim 76; Page 424-431; 443pp; English.
 PS
 XX The present invention describes the human Zmax1 gene and the high bone
 CC mass (HBM) gene, which are found on chromosome 11q13.3. The Zmax1 and
 CC HBM genes have osteopathic activities. The genes can be used in gene
 CC therapy, antisense therapy and in the production of vaccines. They
 CC can be used in the diagnosis and treatment of bone disorders including
 CC osteoporosis, Paget's disease, sclerostosis, osteomalacia and fibrous
 CC dysplasia. ABA82038 to ABA82700 and AAG68168 to AAG68193 represent
 CC sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 2861 AA;
 Query Match 73.8%; Score 31; DB 22; Length 2861;

Best Local Similarity 66.7%; Pred. No. 8.7e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
| | | : : | | |
Db 225 QRIQSESF 233

RESULT 26
ABG06312
ID ABG06312 standard; Protein; 2883 AA.
XX
AC ABG06312;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6303.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS70499.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 36671; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2883 AA;
Query Match 73.8%; Score 31; DB 22; Length 2883;
Best Local Similarity 66.7%; Pred. No. 8.8e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

Best Local Similarity 66.7%; Pred. No. 8.7e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9
| | | : : | | |
Db 247 QRIQSESF 255

RESULT 27
ABG20756
ID ABG20756 standard; Protein; 2980 AA.
XX
AC ABG20756;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #20747.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS84943.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity -
XX
PS Claim 20; SEQ ID No 51115; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 2980 AA;
Query Match 73.8%; Score 31; DB 22; Length 2980;
Best Local Similarity 66.7%; Pred. No. 9.1e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKSF 9

DB 173 QRIQSSSEF 181
||||:|:|

RESULT 28
ABG17025
ID ABG17025 standard; Protein: 3038 AA.
XX
AC ABG17025;
XX
DT 18-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #17016.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS81212.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
XX
PS Claim 20; SEQ ID No 47384; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG00010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Sequence 3038 AA;
Query Match 73.8%; Score 31; DB 22; Length 3038;
Best Local Similarity 66.7%; Pred. NO. 9.3e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 QRIKASKSF 9
||||:|:|
DB 225 QRIQSSSEF 233

RESULT 29
ABG06311
ID ABG06311 standard; Protein: 3060 AA.
XX
AC ABG06311;
XX
DT 13-FEB-2002 (first entry)
XX
DE Novel human diagnostic protein #6302.
XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
food supplement; medical imaging; diagnostic; genetic disorder.
XX
OS Homo sapiens.
XX
PN WO200175067-A2.
XX
PD 11-OCT-2001.
XX
PF 30-MAR-2001; 2001WO-US08631.
XX
PR 31-MAR-2000; 2000US-0540217.
XX
PR 23-AUG-2000; 2000US-0649167.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Drmanac RT, Liu C, Tang YT;
XX
DR WPI: 2001-639362/73.
DR N-PSDB; AAS70498.
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits and to assess
biodiversity -
XX
PS Claim 20; SEQ ID No 36670; 103pp; English.
XX
CC The invention relates to isolated polynucleotide (I) and
polypeptide (II) sequences. (I) is useful as hybridisation probes,
polymerase chain reaction (PCR) primers, oligomers, and for chromosome
and gene mapping, and in recombinant production of (II). The
polynucleotides are also used in diagnostics as expressed sequence tags
for identifying expressed genes. (I) is useful in gene therapy techniques
to restore normal activity of (II) or to treat disease states involving
(II). (II) is useful for generating antibodies against it, detecting or
quantitating a polypeptide in tissue, as molecular weight markers and as
a food supplement. (II) and its binding partners are useful in medical
imaging of sites expressing (II). (I) and (II) are useful for treating
disorders involving aberrant protein expression or biological activity.
The polypeptide and polynucleotide sequences have applications in
diagnostics, forensics, gene mapping, identification of mutations
responsible for genetic disorders or other traits to assess biodiversity
and to produce other types of data and products dependent on DNA and
amino acid sequences. ABG00010-ABG30377 represent novel human
diagnostic amino acid sequences of the invention.
CC Note: The sequence data for this patent did not appear in the printed
specification, but was obtained in electronic format directly from WIPO
at ftp.wipo.int/pub/published_pct_sequences.

Sequence 3060 AA;
Query Match 73.8%; Score 31; DB 22; Length 3060;
Best Local Similarity 66.7%; Pred. NO. 9.4e+02;
Matches 6; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
QY 1 QRIKASKSF 9
||||:|:|
DB 247 QRIQSSSEF 255

RESULT 30
ID AAY99046 standard; Peptide; 18 AA.

AC AAY99046;
XX 07-AUG-2000 (first entry)

DE HLA class II binding antigen epitope peptide #235.

XX Human leucocyte antigen; HLA class II; antigen epitope; pharmaceutical;
KW immune response; chronic viral disease; cancer; autoimmune disease;
KW rheumatoid arthritis; multiple sclerosis; myasthenia gravis; AIDS;
KW allograft rejection; allergy; Lyme disease; hepatitis; prostate cancer;
KW glomerulonephritis; food hypersensitivity; malaria.

XX Unidentified.

XX WO9961916-A1.

XX 02-DEC-1999.

XX 28-MAY-1999; 99WO-US12066.

XX 29-MAY-1998; 98US-0087192.

XX (EPIM-) EPIMMUNE INC.

XX Sette A, Southwood S, Sidney J;

XX WPI; 2000-097143/08.

XX New compositions containing immunogenic peptide epitopes for various
PT HLA class II DR molecules useful for inducing helper T cell response

XX Claim 1; Page 43; 60pp; English.

CC The present invention relates to a new pharmaceutical composition
CC comprising a unit dose form of a peptide, or analogue, comprising an
CC epitope selected from those represented by peptides AAY98612-Y98339
CC which are derived from various antigens for various human leucocyte
CC antigen class DR molecules, representative of the world wide population.
CC The peptide/analogue binds to an HLA class II molecule at an IC-50 of
CC less than or equal to 1,000 nM. The pharmaceutical can be used to induce
CC a helper T cell response. The pharmaceutical focuses the immune response
CC towards selected determinants and could therefore be used in cases of
CC chronic viral diseases and cancer. Examples of diseases that can be
CC treated using the peptide containing pharmaceutical include autoimmune
CC diseases (rheumatoid arthritis, multiple sclerosis, and myasthenia
CC gravis), allograft rejection, allergies, Lyme disease, hepatitis,
CC post-streptococcal endocarditis or glomerulonephritis and food
CC hypersensitivities. The peptide epitopes can be used to enhance immune
CC responses against other immunogens administered with the peptides.
CC Diseases which can be treated using immunogenic mixtures include
CC prostate cancer, hepatitis B, hepatitis C, AIDS, renal carcinoma,
CC cervical carcinoma, lymphoma, and condyloma acuminatum. The peptides may
CC also be used to make monoclonal antibodies useful as potential diagnostic
CC or therapeutic agents. The peptides may also be useful as diagnostic
CC reagents, for example, to determine the susceptibility of an individual
CC to a treatment regimen. Also, the peptides may be used to predict which
CC individuals will be at substantial risk of developing chronic infection.
CC The selection of appropriate T and B cell epitopes should allow the
CC development of epitope based vaccines particularly towards conserved
CC epitopes of pathogens which are characterized by high sequence
CC variability such as HIV, HCV and Malaria.

XX Sequence 18 AA;

Query Match 71.4%; Score 30; DB 21; Length 18;
Best Local Similarity 55.6%; Pred. No. 7.2;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

OY 1 ORIKASKSF 9

Db	1:: :	5	OKVKANSSF 13
RESULT 31			
AAU20688			
ID	AAU20688 standard; Protein; 71 AA.		
XX			
AC	AAU20688;		
XX			
DT	17-DEC-2001 (first entry)		
DE	Human novel foetal antigen, SEQ ID NO 932.		
XX			
KW	Human: foetal tissue antigen; antiinflammatory; neuroprotective;		
KW	immunomodulator; cardiovascular; cytostatic; nephrotoxic;		
KW	cardiovascular; autoimmune disease; rheumatoid arthritis;		
KW	hyperproliferative disorder; breast neoplasm; cancer;		
KW	cardiovascular disorder; cardiac arrest; cerebrovascular disorder;		
KW	cerebral ischaemia; angiogenesis; nervous system disorder;		
KW	Alzheimer's disease; infection; ocular disorder; corneal infection;		
KW	wound healing; epithelial cell proliferation; food additive.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200155312-A2.		
XX			
PD	02-AUG-2001.		
XX			
PF	17-JAN-2001; 2001WO-US01321.		
XX			
PR	31-JAN-2000; 2000US-0179065.		
PR	04-FEB-2000; 2000US-0180628.		
PR	24-FEB-2000; 2000US-0184664.		
PR	02-MAR-2000; 2000US-0186350.		
PR	16-MAR-2000; 2000US-0189874.		
PR	17-MAR-2000; 2000US-0190076.		
PR	18-APR-2000; 2000US-0198123.		
PR	19-MAY-2000; 2000US-0205515.		
PR	07-JUN-2000; 2000US-0209467.		
PR	28-JUN-2000; 2000US-0214886.		
PR	30-JUN-2000; 2000US-0215135.		
PR	07-JUL-2000; 2000US-0216647.		
PR	11-JUL-2000; 2000US-0216880.		
PR	11-JUL-2000; 2000US-0217487.		
PR	14-JUL-2000; 2000US-0217496.		
PR	26-JUL-2000; 2000US-0218290.		
PR	26-JUL-2000; 2000US-0220963.		
PR	26-JUL-2000; 2000US-0220964.		
PR	14-AUG-2000; 2000US-0224518.		
PR	14-AUG-2000; 2000US-0224519.		
PR	14-AUG-2000; 2000US-0225213.		
PR	14-AUG-2000; 2000US-0225214.		
PR	14-AUG-2000; 2000US-0225217.		
PR	14-AUG-2000; 2000US-0225447.		
PR	14-AUG-2000; 2000US-0225757.		
PR	14-AUG-2000; 2000US-0225758.		
PR	14-AUG-2000; 2000US-0225759.		
PR	18-AUG-2000; 2000US-0226279.		
PR	22-AUG-2000; 2000US-0226681.		
PR	22-AUG-2000; 2000US-0226686.		
PR	22-AUG-2000; 2000US-0227182.		
PR	23-AUG-2000; 2000US-0227009.		
PR	30-SEP-2000; 2000US-0228924.		
PR	01-SEP-2000; 2000US-0229287.		
PR	01-SEP-2000; 2000US-0229343.		
PR	01-SEP-2000; 2000US-0229344.		
PR	01-SEP-2000; 2000US-0229345.		
PR	05-SEP-2000; 2000US-0229509.		
PR	05-SEP-2000; 2000US-0229513.		

PR	06-SEP-2000;	2000US-0230437;
PR	06-SEP-2000;	2000US-0230438;
PR	08-SEP-2000;	2000US-0231242;
PR	08-SEP-2000;	2000US-0231243;
PR	08-SEP-2000;	2000US-0231244;
PR	08-SEP-2000;	2000US-0231413;
PR	08-SEP-2000;	2000US-0231414;
PR	08-SEP-2000;	2000US-0232080;
PR	08-SEP-2000;	2000US-0232081;
PR	12-SEP-2000;	2000US-0231968;
PR	14-SEP-2000;	2000US-0232397;
PR	14-SEP-2000;	2000US-0232398;
PR	14-SEP-2000;	2000US-0232399;
PR	14-SEP-2000;	2000US-0232400;
PR	14-SEP-2000;	2000US-0232401;
PR	14-SEP-2000;	2000US-0233063;
PR	14-SEP-2000;	2000US-0233064;
PR	14-SEP-2000;	2000US-0233065;
PR	21-SEP-2000;	2000US-0234222;
PR	21-SEP-2000;	2000US-0234277;
PR	25-SEP-2000;	2000US-0234997;
PR	25-SEP-2000;	2000US-0234998;
PR	26-SEP-2000;	2000US-0235484;
PR	27-SEP-2000;	2000US-0235834;
PR	27-SEP-2000;	2000US-0235836;
PR	29-SEP-2000;	2000US-0236327;
PR	29-SEP-2000;	2000US-0236367;
PR	29-SEP-2000;	2000US-0236368;
PR	29-SEP-2000;	2000US-0236369;
PR	29-SEP-2000;	2000US-0236370;
PR	02-OCT-2000;	2000US-0236602;
PR	02-OCT-2000;	2000US-0237037;
PR	02-OCT-2000;	2000US-0237038;
PR	02-OCT-2000;	2000US-0237039;
PR	02-OCT-2000;	2000US-0237040;
PR	13-OCT-2000;	2000US-0239935;
PR	13-OCT-2000;	2000US-0239937;
PR	20-OCT-2000;	2000US-0240960;
PR	20-OCT-2000;	2000US-0241221;
PR	20-OCT-2000;	2000US-0241785;
PR	20-OCT-2000;	2000US-0241786;
PR	20-OCT-2000;	2000US-0241787;
PR	20-OCT-2000;	2000US-0241808;
PR	20-OCT-2000;	2000US-0241809;
PR	20-OCT-2000;	2000US-0241826;
PR	01-NOV-2000;	2000US-0244617;
PR	08-NOV-2000;	2000US-0246474;
PR	08-NOV-2000;	2000US-0246475;
PR	08-NOV-2000;	2000US-0246476;
PR	08-NOV-2000;	2000US-0246477;
PR	08-NOV-2000;	2000US-0246478;
PR	08-NOV-2000;	2000US-0245523;
PR	08-NOV-2000;	2000US-0245524;
PR	08-NOV-2000;	2000US-0245525;
PR	08-NOV-2000;	2000US-0245526;
PR	08-NOV-2000;	2000US-0245527;
PR	08-NOV-2000;	2000US-0245528;
PR	08-NOV-2000;	2000US-0245532;
PR	08-NOV-2000;	2000US-0246609;
PR	08-NOV-2000;	2000US-0246610;
PR	08-NOV-2000;	2000US-0246611;
PR	08-NOV-2000;	2000US-0246613;
PR	17-NOV-2000;	2000US-0249207;
PR	17-NOV-2000;	2000US-0249208;
PR	17-NOV-2000;	2000US-0249209;
PR	17-NOV-2000;	2000US-0249210;
PR	17-NOV-2000;	2000US-0249211;
PR	17-NOV-2000;	2000US-0249212;
PR	17-NOV-2000;	2000US-0249213;
PR	17-NOV-2000;	2000US-0249214;
PR	17-NOV-2000;	2000US-0249215;
PR	17-NOV-2000;	2000US-0249216;
PR	17-NOV-2000;	2000US-0249217;

PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	03-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	03-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX	Rosen CA, Barash SC, Ruben SM;	
PI	WPI; 2001-488782/53.	
DR	N-PSDB; AAS33508.	
XX	New polynucleotides and polypeptides for diagnosing, treating,	
PT	preventing or prognosing e.g. diseases or disorders of the nervous,	
PT	musculoskeletal, excretory, gastrointestinal, reproductive, and	
PT	respiratory systems	
XX	Claim 11; SEQ ID NO 932; 642pp; English.	
XX	The invention relates to novel nucleic acids encoding novel human foetal	
CC	antigens. The nucleic acids and proteins are used to prevent, treat (e.g.	
CC	by gene therapy) or ameliorate a medical condition in e.g. humans, mice,	
CC	rabbits, goats, horses, cats, dogs, chickens or sheep. They	
CC	are also used in diagnosing a pathological condition or susceptibility	
CC	to a pathological condition. The antibodies to the antigens can also	
CC	be used in alleviating symptoms associated with the disorders and in	
CC	diagnostic immunoassays e.g. radioimmunoassays or enzyme linked	
CC	immunosorbent assays (ELISA). Disorders which are diagnosed or treated	
CC	include autoimmune diseases e.g. rheumatoid arthritis,	
CC	hyperproliferative disorders e.g. neoplasms of the breast or liver,	
CC	cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders	
CC	e.g. cerebral ischaemia, anglogenesis, nervous system disorders e.g.	
CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi	
CC	and ocular disorders e.g. corneal infection. The polypeptides can also	
CC	be used to aid wound healing and epithelial cell proliferation, to	
CC	prevent skin aging due to sunburn, to maintain organs before	
CC	transplantation, for supporting cell culture of primary tissues, to	
CC	regenerate tissues and in chemotaxis. The polypeptides can also be used	
CC	as a food additive or preservative to increase or decrease storage	
CC	capabilities, fat content, lipid, protein, carbohydrate, vitamins,	
CC	minerals, cofactors and other nutritional components. Numerous	
CC	examples of diseases and disorders treated by the nucleic acids and	
CC	proteins are given in the specification. The present sequence	
Query Match 71.4% Score 30; DB 22; Length 71;		

Qy 2 RIKASKSF 9
| : | | | |
Db 56 RLKMSKSF 6

RESULT 32
AAG20817
ID AAG20817 standard: protein: 343 AA:

XX AAG20817; PR 18-JUN-1999; 99US-0139462.
XX AC PR 18-JUN-1999; 99US-0139463.
XX DT PR 18-JUN-1999; 99US-0139750.
XX XX PR 18-JUN-1999; 99US-0139763.
DE 17-OCT-2000 (first entry) PR 21-JUN-1999; 99US-0139817.
XX Arabidopsis thaliana protein fragment SEQ ID NO: 23147. PR 23-JUN-1999; 99US-0139899.
XX KW Protein identification; signal transduction pathway; metabolic pathway; PR 23-JUN-1999; 99US-0140353.
KW hybridisation assay; genetic mapping; gene expression control; promoter; PR 24-JUN-1999; 99US-0140695.
KW termination sequence. PR 28-JUN-1999; 99US-0140823.
XX XX PR 29-JUN-1999; 99US-0140991.
OS Arabidopsis thaliana. PR 30-JUN-1999; 99US-0141287.
XX XX PR 01-JUL-1999; 99US-0141842.
PN EP1033405-A2. PR 01-JUL-1999; 99US-0142154.
XX XX PR 02-JUL-1999; 99US-0142055.
PD 06-SEP-2000. PR 06-JUL-1999; 99US-0142390.
XX XX PR 08-JUL-1999; 99US-0142803.
PF 25-FEB-2000; 2000EP-0301439. PR 09-JUL-1999; 99US-0142920.
XX XX PR 12-JUL-1999; 99US-0142977.
XX 25-FEB-1999; 99US-0121825. PR 13-JUL-1999; 99US-0143542.
XX 05-MAR-1999; 99US-0123180. PR 14-JUL-1999; 99US-0143624.
PR 09-MAR-1999; 99US-0123548. PR 15-JUL-1999; 99US-0144005.
PR 23-MAR-1999; 99US-0125788. PR 16-JUL-1999; 99US-0144086.
PR 29-MAR-1999; 99US-0126264. PR 16-JUL-1999; 99US-0144086.
PR 25-MAR-1999; 99US-0126785. PR 19-JUL-1999; 99US-0144331.
PR 01-APR-1999; 99US-0127462. PR 19-JUL-1999; 99US-0144332.
PR 06-APR-1999; 99US-0128234. PR 19-JUL-1999; 99US-0144333.
PR 08-APR-1999; 99US-0128714. PR 19-JUL-1999; 99US-0144334.
PR 16-APR-1999; 99US-0129845. PR 19-JUL-1999; 99US-0144335.
PR 19-APR-1999; 99US-0130077. PR 20-JUL-1999; 99US-0144352.
PR 21-APR-1999; 99US-0130449. PR 20-JUL-1999; 99US-0144632.
PR 23-APR-1999; 99US-0130510. PR 20-JUL-1999; 99US-0144884.
PR 23-APR-1999; 99US-0130891. PR 21-JUL-1999; 99US-0144814.
PR 28-APR-1999; 99US-0131449. PR 21-JUL-1999; 99US-0145086.
PR 30-APR-1999; 99US-0132048. PR 21-JUL-1999; 99US-0145088.
PR 30-APR-1999; 99US-0132407. PR 22-JUL-1999; 99US-0145087.
PR 04-MAY-1999; 99US-0132484. PR 22-JUL-1999; 99US-0145089.
PR 05-MAY-1999; 99US-0132485. PR 22-JUL-1999; 99US-0145192.
PR 06-MAY-1999; 99US-0132486. PR 22-JUL-1999; 99US-0145192.
PR 06-MAY-1999; 99US-0132487. PR 23-JUL-1999; 99US-0145145.
PR 07-MAY-1999; 99US-0132863. PR 23-JUL-1999; 99US-0145218.
PR 11-MAY-1999; 99US-0134256. PR 23-JUL-1999; 99US-0145224.
PR 14-MAY-1999; 99US-0134218. PR 26-JUL-1999; 99US-0145276.
PR 14-MAY-1999; 99US-0134219. PR 27-JUL-1999; 99US-0145913.
PR 14-MAY-1999; 99US-0134321. PR 27-JUL-1999; 99US-0145918.
PR 18-MAY-1999; 99US-0134370. PR 27-JUL-1999; 99US-0145919.
PR 18-MAY-1999; 99US-0134768. PR 27-JUL-1999; 99US-0145919.
PR 19-MAY-1999; 99US-0134941. PR 28-JUL-1999; 99US-0145951.
PR 20-MAY-1999; 99US-0135124. PR 02-AUG-1999; 99US-0146386.
PR 21-MAY-1999; 99US-0135353. PR 02-AUG-1999; 99US-0146388.
PR 24-MAY-1999; 99US-0135629. PR 02-AUG-1999; 99US-0146389.
PR 25-MAY-1999; 99US-0136021. PR 03-AUG-1999; 99US-0147038.
PR 27-MAY-1999; 99US-0136392. PR 04-AUG-1999; 99US-0147204.
PR 28-MAY-1999; 99US-0136782. PR 04-AUG-1999; 99US-0147302.
PR 01-JUN-1999; 99US-0137222. PR 05-AUG-1999; 99US-0147192.
PR 03-JUN-1999; 99US-0137528. PR 05-AUG-1999; 99US-0147260.
PR 04-JUN-1999; 99US-0137502. PR 06-AUG-1999; 99US-0147303.
PR 07-JUN-1999; 99US-0137724. PR 06-AUG-1999; 99US-0147416.
PR 08-JUN-1999; 99US-0138094. PR 09-AUG-1999; 99US-0147493.
PR 10-JUN-1999; 99US-0138540. PR 09-AUG-1999; 99US-0147935.
PR 10-JUN-1999; 99US-0138847. PR 10-AUG-1999; 99US-0148171.
PR 14-JUN-1999; 99US-0139119. PR 11-AUG-1999; 99US-0148319.
PR 16-JUN-1999; 99US-0139452. PR 12-AUG-1999; 99US-0148341.
PR 16-JUN-1999; 99US-0139453. PR 13-AUG-1999; 99US-0148565.
PR 17-JUN-1999; 99US-0139492. PR 13-AUG-1999; 99US-0148684.
PR 18-JUN-1999; 99US-0139454. PR 16-AUG-1999; 99US-0149368.
PR 18-JUN-1999; 99US-0139455. PR 17-AUG-1999; 99US-0149175.
PR 18-JUN-1999; 99US-0139456. PR 18-AUG-1999; 99US-0149426.
PR 18-JUN-1999; 99US-0139457. PR 20-AUG-1999; 99US-0149722.
PR 18-JUN-1999; 99US-0139458. PR 20-AUG-1999; 99US-0149723.
PR 18-JUN-1999; 99US-0139459. PR 20-AUG-1999; 99US-0149929.
PR 18-JUN-1999; 99US-0139460. PR 23-AUG-1999; 99US-0149902.
PR 18-JUN-1999; 99US-0139461. PR 23-AUG-1999; 99US-0149930.


```
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 18-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 71.4%; Score 30; DB 21; Length 343;
Best Local Similarity 66.7%; Pred. No. 1.6e+02;
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9
Db 56 RRVKASVSF 64

RESULT 33
AAG39402
ID AAG39402 standard; Protein; 351 AA.
XX
AC AAG39402;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 48747.
XX
```

```
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
XX 06-SEP-2000.
XX
PD 25-FEB-2000; 2000EP-0301439.
PF
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
```

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

23-APR-1999; 99US-0130891.

28-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

30-APR-1999; 99US-0132407.

04-MAY-1999; 99US-0132484.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

06-MAY-1999; 99US-0132487.

07-MAY-1999; 99US-0132863.

11-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

14-MAY-1999; 99US-0134370.

18-MAY-1999; 99US-0134768.

19-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.

Arabidopsis thaliana.

EP1033405-A2.

06-SEP-2000.

25-FEB-2000; 2000EP-0301439.

25-FEB-1999; 99US-0121825.

05-MAR-1999; 99US-0123180.

09-MAR-1999; 99US-0123548.

23-MAR-1999; 99US-0125788.

25-MAR-1999; 99US-0126264.

29-MAR-1999; 99US-0126785.

01-APR-1999; 99US-0127462.

06-APR-1999; 99US-0128234.

08-APR-1999; 99US-0128714.

16-APR-1999; 99US-0129845.

19-APR-1999; 99US-0130077.

21-APR-1999; 99US-0130449.

23-APR-1999; 99US-0130510.

23-APR-1999; 99US-0130891.

28-APR-1999; 99US-0131449.

30-APR-1999; 99US-0132048.

30-APR-1999; 99US-0132407.

04-MAY-1999; 99US-0132484.

05-MAY-1999; 99US-0132485.

06-MAY-1999; 99US-0132486.

06-MAY-1999; 99US-0132487.

07-MAY-1999; 99US-0132863.

11-MAY-1999; 99US-0134256.

14-MAY-1999; 99US-0134218.

14-MAY-1999; 99US-0134219.

14-MAY-1999; 99US-0134221.

14-MAY-1999; 99US-0134370.

18-MAY-1999; 99US-0134768.

19-MAY-1999; 99US-0134941.

20-MAY-1999; 99US-0135124.

21-MAY-1999; 99US-0135353.

24-MAY-1999; 99US-0135629.

25-MAY-1999; 99US-0136021.

27-MAY-1999; 99US-0136392.

PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141387.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149502.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.

PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160880.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 71.4%; Score 30; DB 21; Length 351;
Best Local Similarity 75.0%; Pred. No. 1.6e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 2 RIKASKSF 9
Db 14 RIKVSKTF 21
||| ||:|

RESULT 34

AAG20816
ID AAG20816 standard; Protein: 427 AA.
XX
AC AAG20816;
XX

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 23146.

DE Arabidopsis thaliana protein fragment SEQ ID NO: 23146.
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX Arabidopsis thaliana.
PN EP1033405-A2.

PR	22-SEP-1999;	99US-0155139.	PR	09-MAR-1999;	99US-0123548.
PR	23-SEP-1999;	99US-0155486.	PR	23-MAR-1999;	99US-0125788.
PR	24-SEP-1999;	99US-0155659.	PR	25-MAR-1999;	99US-0126264.
PR	28-SEP-1999;	99US-0156458.	PR	29-MAR-1999;	99US-0125785.
PR	29-SEP-1999;	99US-0156596.	PR	01-APR-1999;	99US-0127462.
PR	04-OCT-1999;	99US-0157117.	PR	06-APR-1999;	99US-0128234.
PR	05-OCT-1999;	99US-0157753.	PR	08-APR-1999;	99US-0128714.
PR	06-OCT-1999;	99US-0157865.	PR	16-APR-1999;	99US-0129845.
PR	07-OCT-1999;	99US-0158029.	PR	19-APR-1999;	99US-0130077.
PR	08-OCT-1999;	99US-0158232.	PR	21-APR-1999;	99US-0130449.
PR	12-OCT-1999;	99US-0158369.	PR	23-APR-1999;	99US-0130510.
PR	13-OCT-1999;	99US-0159293.	PR	23-APR-1999;	99US-0130891.
PR	13-OCT-1999;	99US-0159294.	PR	28-APR-1999;	99US-0131449.
PR	14-OCT-1999;	99US-0159295.	PR	30-APR-1999;	99US-0132048.
PR	14-OCT-1999;	99US-0159329.	PR	30-APR-1999;	99US-0132407.
PR	14-OCT-1999;	99US-0159330.	PR	04-MAY-1999;	99US-0132484.
PR	14-OCT-1999;	99US-0159331.	PR	05-MAY-1999;	99US-0132485.
PR	14-OCT-1999;	99US-0159637.	PR	06-MAY-1999;	99US-0132486.
PR	14-OCT-1999;	99US-0159638.	PR	06-MAY-1999;	99US-0132487.
PR	18-OCT-1999;	99US-0159584.	PR	07-MAY-1999;	99US-0132863.
PR	21-OCT-1999;	99US-0160741.	PR	11-MAY-1999;	99US-0134256.
PR	21-OCT-1999;	99US-0160767.	PR	14-MAY-1999;	99US-0134218.
PR	21-OCT-1999;	99US-0160768.	PR	14-MAY-1999;	99US-0134219.
PR	21-OCT-1999;	99US-0160770.	PR	14-MAY-1999;	99US-0134221.
PR	21-OCT-1999;	99US-0160814.	PR	14-MAY-1999;	99US-0134370.
PR	21-OCT-1999;	99US-0160815.	PR	18-MAY-1999;	99US-0134768.
PR	22-OCT-1999;	99US-0160980.	PR	19-MAY-1999;	99US-0134941.
PR	22-OCT-1999;	99US-0160981.	PR	20-MAY-1999;	99US-0135124.
PR	22-OCT-1999;	99US-0160989.	PR	21-MAY-1999;	99US-0135353.
PR	25-OCT-1999;	99US-0161404.	PR	24-MAY-1999;	99US-0135629.
PR	25-OCT-1999;	99US-0161405.	PR	25-MAY-1999;	99US-0136021.
PR	25-OCT-1999;	99US-0161406.	PR	27-MAY-1999;	99US-0136392.
PR	26-OCT-1999;	99US-0161359.	PR	28-MAY-1999;	99US-0136782.
PR	26-OCT-1999;	99US-0161360.	PR	01-JUN-1999;	99US-0137222.
PR	26-OCT-1999;	99US-0161361.	PR	03-JUN-1999;	99US-0137528.
PR	28-OCT-1999;	99US-0161320.	PR	04-JUN-1999;	99US-0137502.
PR	28-OCT-1999;	99US-0161992.	PR	07-JUN-1999;	99US-0137724.
PR	28-OCT-1999;	99US-0161993.	PR	08-JUN-1999;	99US-0138094.
PR	28-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138540.
PR	29-OCT-1999;	99US-0162142.	PR	10-JUN-1999;	99US-0138847.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUN-1999;	99US-0139119.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139452.
PR	29-OCT-1999;	99US-0162142.	PR	16-JUN-1999;	99US-0139453.
PR	29-OCT-1999;	99US-0162142.	PR	17-JUN-1999;	99US-0139492.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139454.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139455.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139456.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139457.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139458.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139459.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139460.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139461.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139462.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139463.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139750.
PR	29-OCT-1999;	99US-0162142.	PR	18-JUN-1999;	99US-0139763.
PR	29-OCT-1999;	99US-0162142.	PR	21-JUN-1999;	99US-0139817.
PR	29-OCT-1999;	99US-0162142.	PR	22-JUN-1999;	99US-0139899.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140353.
PR	29-OCT-1999;	99US-0162142.	PR	23-JUN-1999;	99US-0140354.
PR	29-OCT-1999;	99US-0162142.	PR	24-JUN-1999;	99US-0140695.
PR	29-OCT-1999;	99US-0162142.	PR	28-JUN-1999;	99US-0140823.
PR	29-OCT-1999;	99US-0162142.	PR	29-JUN-1999;	99US-0140991.
PR	29-OCT-1999;	99US-0162142.	PR	30-JUN-1999;	99US-0141287.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0141842.
PR	29-OCT-1999;	99US-0162142.	PR	01-JUL-1999;	99US-0142154.
PR	29-OCT-1999;	99US-0162142.	PR	02-JUL-1999;	99US-0142055.
PR	29-OCT-1999;	99US-0162142.	PR	06-JUL-1999;	99US-0142390.
PR	29-OCT-1999;	99US-0162142.	PR	08-JUL-1999;	99US-0142803.
PR	29-OCT-1999;	99US-0162142.	PR	09-JUL-1999;	99US-0142920.
PR	29-OCT-1999;	99US-0162142.	PR	12-JUL-1999;	99US-0142977.
PR	29-OCT-1999;	99US-0162142.	PR	13-JUL-1999;	99US-0143542.
PR	29-OCT-1999;	99US-0162142.	PR	14-JUL-1999;	99US-0143624.

Query Match 71.4%; Score 30; DB 21; Length 427;

Best Local Similarity 66.7%; Pred. No. 2e+02;

Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 QRIKAKSF 9

DB 140 RRVKASVSF 148

RESULT 35
AAG20815
ID AAG20815 standard; Protein; 448 AA.

AC AAG20815;
XX
XX DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 23145.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

XX

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.

PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 14-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 71.4%; Score 30; DB 21; Length 448;
Best Local Similarity 66.7%; Pred. No. 2.le+02;
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 ORIKASKSF 9
:I:III II
Db 161 RRVKASVSF 169

RESULT 36

AAG39401

ID AAG39401 standard; Protein; 605 AA.

XX AAG39401;

AC AAG39401;

XX 18-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 48746.

DE Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

XX EP1033405-A2.

PN 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140895.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142154.
PR 05-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151438.
PR 07-SEP-1999; 99US-0151930.
PR 10-SEP-1999; 99US-0152363.
PR 13-SEP-1999; 99US-0153070.
PR 15-SEP-1999; 99US-0153758.
PR 16-SEP-1999; 99US-0154018.
PR 20-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 71.4%; Score 30; DB 21; Length 605;

Best Local Similarity 75.0%; Pred. No. 2.9e+02; Mismatches 1; Indels 0; Gaps 0;

Matches 6; Conservative 1;

QY 2 RIKASKSF 9

||| ||:|

Db 268 RIKVSKTF 275

RESULT 37

AAB36463

ID AAB36463 standard; Protein; 726 AA.

AC AAB36463;

DT 01-MAR-2001 (first entry)

DE Human plakophilin ppla (PKP1a) protein SEQ ID NO:3.

KW Plakophilin-3; PKP3; catenin-like protein; characterisation; diagnosis;
 KW desmosome; epithelial cell; skin disease; dermatological; gene therapy;
 KW vaccine.

OS Homo sapiens.

PN WO200066619-A2.

PD 09-NOV-2000.

PF 28-APR-2000; 2000WO-EP04389.

PR 30-APR-1999; 99EP-0870093.

PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.

PI Van Roy F, Bonne S;

DR WPI; 2000-687529/67.

XX Nucleic acids encoding Plakophilin-3 polypeptides, useful for treating
 PT skin diseases and disorders of epithelial tissue associated with
 PT inappropriate Plakophilin-3 expression and activity -

XX Example 1; Fig 4; 132pp; English.

XX The present invention describes an isolated or recombinant nucleic acid
 CC molecule (I) encoding a Plakophilin-3 (PKP3), from humans, mice and

CC Xenopus laevis. (I) has dermatological activity, and can be used in
 CC gene therapy and for vaccines. (I) and the protein it encodes may be
 CC used in the prevention, treatment and diagnosis of diseases associated
 CC with inappropriate PKP3 expression, such as skin diseases and disorders
 CC affecting epithelial tissue. For example, (I) (and vectors containing
 CC (I)) and the PKP3 polypeptide may be used to treat disorders associated
 CC with decreased PKP3 expression by rectifying mutations or deletions in
 CC a patient's genome that affect the activity of PKP3 by expressing
 CC inactive proteins or to supplement the patients own production of PKP3
 CC polypeptides. Additionally, (I) may be used to produce PKP3, according
 CC to standard recombinant DNA methodology, by inserting the nucleic acids
 CC into a host cell and culturing the cell to express the protein. (I) and
 CC complementary sequences may also be used as DNA probes in diagnostic
 CC assays to detect and quantitate the presence of similar nucleic acid
 CC sequences in samples, and hence which patients may be in need of
 CC restorative therapy. The PKP3 polypeptides may be used as antigens in
 CC the production of antibodies against PKP3 and in assays to identify
 CC modulators (agonists and antagonists) of PKP3 expression and activity.
 CC The anti-PKP3 antibodies and PKP3 antagonists may also be used to down
 CC regulate PKP3 expression and activity. The anti-PKP3 antibodies may also
 CC be used as diagnostic agents for detecting the presence of PKP3
 CC polypeptides in samples (e.g. by enzyme linked immunosorbant assay
 CC (ELISA)). PKP3 is a catenin-like protein, which is present in desmosomes
 CC and nuclei of epithelial cells. The present sequence represents human
 CC PKP1a (ppla), which is used in the exemplification of the present
 CC invention.

XX Sequence 726 AA;

Query Match 71.4%; Score 30; DB 21; Length 726;

Best Local Similarity 75.0%; Pred. No. 3.5e+02;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 QRIKASKS 8

|:||||:|

Db 148 QRIKASRS 155

RESULT 38

AAM94095

ID AAM94095 standard; Protein; 46 AA.

AC AAM94095;

XX 21-NOV-2001 (first entry)

DE Human reproductive system related antigen SEQ ID NO: 2753.

KW Human; reproductive system related antigen; reproductive system disorder;
 KW cancer; gene therapy.

OS Homo sapiens.

PN WO200155320-A2.

PD 02-AUG-2001.

PF 17-JAN-2001; 2001WO-US01339.

PR 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240360.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465570/50.
XX N-PSDB; AAL00065.
PT Isolated nucleic acid molecule encoding a reproductive system antigen -
PT is used in preventing, treating or ameliorating a medical condition -
PS Claim 11; SEQ ID NO 2753; 1297pp + Sequence Listing; English.
XX The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention.
XX Sequence 46 AA;

Query Match 69.0%; Score 29; DB 22; Length 46;
Best Local Similarity 62.5%; Pred. No. 32;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

	Best Local Similarity	55.6%;	Pred. NO. 57;	
	Matches	5; Conservative	3; Mismatches	1; Indels
	Gaps	0;		

QY	1 QRIKASKSF 9	
Db	70 QRIRAGRSY 78	

RESULT 39	
ID	AAM41295 standard; Protein; 79 AA.
AC	AAM41295;
DT	22-OCT-2001 (first entry)
XX	Human polypeptide SEQ ID NO 6226.
DE	Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW	peripheral nervous system; neuropathy; central nervous system; CNS;
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW	chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW	leukaemia.
OS	Homo sapiens.
XX	
PN	WO200153312-A1.
XX	
PD	26-JUL-2001.
XX	
PF	26-DEC-2000; 2000WO-US34263.
XX	
PR	21-JAN-2000; 2000US-0488725.
PR	25-APR-2000; 2000US-0552317.
PR	09-JUL-2000; 2000US-0598042.
PR	19-JUL-2000; 2000US-0620312.
PR	03-AUG-2000; 2000US-0653450.
PR	14-SEP-2000; 2000US-0662191.
PR	19-OCT-2000; 2000US-0693036.
PR	29-NOV-2000; 2000US-0727344.
XX	(HYSE-) HYSEQ INC.
PA	
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX	
WI	WPI: 2001-442253/47.
N-PSDB	AAT60451.
XX	
PT	Novel nucleic acids and polypeptides, useful for treating disorders
XX	such as central nervous system injuries -
XX	
PS	Example 2; SEQ ID NO 6226; 10078pp; English.
XX	
CC	The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC	the encoded polypeptides (AAM38642-AAAM42213) with nootropic,
CC	immunosuppressant and cytostatic activity. The polynucleotides are useful
CC	in gene therapy. A composition containing a polypeptide or polynucleotide
CC	of the invention may be used to treat diseases of the peripheral nervous
CC	system, such as peripheral nervous injuries, peripheral neuropathy and
CC	localised neuropathies and central nervous system diseases, such as
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC	utilisation of the activities such as: Immune system suppression,
CC	Activin/inhibin activity, chemotactic/chemokinetic activity, hemostatic
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC	C.N.S disorders.
CC	Note: The sequence data for this patent did not form part of the printed
CC	specification.
XX	
SQ	Sequence 79 AA;

Query Match	69.0%;	Score 29;	DB 22;	Length 79;
Best Local Similarity	55.6%;	Pred. No. 93;		
Matches	5; Conservative	3; Mismatches	1; Indels	0; Gaps
Gaps	0;			

QY	1 QRIKASKSF 9	
Db	107 ERLKSSSSF 115	

RESULT 41	
ID	ABB67070

ID ABB67070 standard; Protein; 182 AA.
XX ABB67070;
AC
XX
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 28002.
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX 23-MAR-2000; 2000US-191637P.
PR
XX
XX 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
XX Venter JC, Adams M, Li PWD, Myers EW;
PI
XX
XX WPI; 2001-656860/75.
DR
XX
XX N-PSDB; ABL11173.
DR
XX
XX New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
XX
XX Disclosure; SEQ ID NO 28002; 21pp + Sequence Listing; English.
PS
XX
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX
XX Sequence 182 AA;
SQ
Query Match 69.0%; Score 29; DB 22; Length 182;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;
QY 1 ORIKASKSF 9
Db :||| |||
61 KRIKLEKSF 69
RESULT 42
AAG82665
ID AAG82665 standard; Protein; 203 AA.
XX
XX AAG82665;
AC
XX
XX 03-SEP-2001 (first entry)
DT
XX
XX S. epidermidis open reading frame protein sequence SEQ ID NO:2424.
DE
XX
XX Staphylococcus epidermidis SR1 strain; infection; diagnosis;
KW vaccination; endocarditis.
XX
XX Staphylococcus epidermidis.
OS
XX

PN WO200134809-A2.
XX
XX 17-MAY-2001.
PD
XX
XX 09-NOV-2000; 2000WO-US30782.
PF
XX
XX 09-NOV-1999; 99US-0164258.
PR
XX
XX (GLAX) GLAXO GROUP LTD.
PA
XX
XX Kimmerly WJ;
PI
XX
XX WPI; 2001-316495/33.
DR
XX
XX N-PSDB; AAH53515.
DR
XX
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PT useful for vaccinating against infections, e.g. endocarditis -
PT
XX
XX Claim 18; Page 639; 2188pp; English.
PS
XX
XX AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC (II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC (I) and (II) can have antibacterial activity and therefore can be used
CC in vaccination. The nucleic acids (I) may be used to produce the
CC S. epidermidis polypeptides (II) via the production of vectors
CC containing them which are used to produce hosts cells which express the
CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC used to vaccinate subjects and to raise antibodies against the bacteria.
CC The polypeptides may also be used to assay for other inhibitors of their
CC activity and therefore identify compounds that may be used for the
CC treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC AAH55090 represent specifically claimed S. epidermidis genomic DNA
CC polynucleotide sequences from the present invention. AAH55091 to
CC AAH55098 represent oligonucleotide sequences and primers which are used
CC in the exemplification of the present invention.
CC N.B. The present invention specifically claims all the polynucleotide
CC sequences given in the sequence listing of the present specification,
CC however the sequence listing only goes up to SEQ ID NO:4454 so even
CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC no sequences are present for SEQ ID NO:4455 to 4464.
XX
XX Sequence 203 AA;
SQ
Query Match 69.0%; Score 29; DB 22; Length 203;
Best Local Similarity 55.6%; Pred. No. 1.5e+02;
Matches 5; Conservative 3; Mismatches 1; Indels 0; Gaps 0;
QY 1 ORIKASKSF 9
Db :||| |||
179 KRLKSKSF 187
RESULT 43
ABB62005
ID ABB62005 standard; Protein; 308 AA.
XX
XX ABB62005;
AC
XX
XX 26-MAR-2002 (first entry)
DT
XX
XX Drosophila melanogaster polypeptide SEQ ID NO 12807.
DE
XX
XX Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
XX Drosophila melanogaster.
OS
XX
XX WO200171042-A2.
PN
XX
XX 27-SEP-2001.
PD
XX
XX 23-MAR-2001; 2001WO-US09231.
PF
XX
XX

PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 PA (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 XX WPI; 2001-656860/75.
 DR N-PSDB; ABL06108.
 XX New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 12807; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX SQ Sequence 308 AA;
 Query Match 69.0%; Score 29; DB 22; Length 308;
 Best Local Similarity 85.7%; Pred. No. 2.4e+02;
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 QRIKASK 7
 Db 171 QRMKASK 177
 II:IIII
 RESULT 44
 AAY91342
 ID AAY91342 standard; Protein; 438 AA.
 AC AAY91342;
 XX 30-MAY-2000 (first entry)
 DT Group B Streptococcus protein sequence SEQ ID NO:75.
 XX Group B Streptococcus agalactiae;
 KW vaccine; screening; immunogen; detection; diagnosis; infection;
 KW antibody; affibody; antibacterial.
 XX Streptococcus agalactiae.
 OS WO200006736-A2.
 PN 10-FEB-2000.
 XX 27-JUL-1999; 99WO-GB02444.
 XX 27-JUL-1998; 98GB-0016335.
 PR 19-MAR-1999; 99US-0125163.
 XX (MICR-) MICROBIAL TECHNIQS LTD.
 PA Le Page RWF, Wells JM, Hanniffy SB;
 PI WPI; 2000-195299/17.
 XX New Group B Streptococcus protein, useful as vaccine, for diagnosis of
 PT Streptococcal infections and for screening of antibodies or affibodies
 XX

PS Claim 1; Fig 1; 123pp; English.
 XX AAA05803 to AAA05872 encode proteins, polypeptides and peptides (given
 CC in AAY91275 to AAY91343) isolated from Group B Streptococcus (GBS), also
 CC known as Streptococcus agalactiae. The GBS polynucleotides and
 CC polypeptides have antibacterial activity. Immunogenic compositions
 CC comprising GBS polynucleotides or polypeptides can be used as vaccines
 CC and for the treatment or prophylaxis of GBS infection. The
 CC polynucleotides and polypeptides can also be used in the detection of GBS
 CC and for screening DNA encoding bacterial cell envelope associated or
 CC secreted antigens in gram positive bacteria. AAA05873 to AAA05941
 CC represent primers used in the exemplification of the present invention.
 XX SQ Sequence 438 AA;
 Query Match 69.0%; Score 29; DB 21; Length 438;
 Best Local Similarity 62.5%; Pred. No. 3.4e+02;
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;
 QY 2 RIKASKSF 9
 Db 333 KVKASKKF 340
 :::IIII I
 RESULT 45
 ABB69682
 ID ABB69682 standard; Protein; 669 AA.
 XX AC ABB69682;
 XX 26-MAR-2002 (first entry)
 DT Drosophila melanogaster polypeptide SEQ ID NO 35838.
 DE Drosophila; developmental biology; cell signalling; insecticide;
 XX pharmaceutical.
 KW Drosophila melanogaster.
 XX OS WO200171042-A2.
 PN 27-SEP-2001.
 PD 23-MAR-2001; 2001WO-US09231.
 PF 23-MAR-2000; 2000US-191637P.
 XX 11-JUL-2000; 2000US-0614150.
 PR (PEKE) PE CORP NY.
 XX Venter JC, Adams M, Li PWD, Myers EW;
 PI WPI; 2001-656860/75.
 XX N-PSDB; ABL13785.
 DR New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX Disclosure; SEQ ID NO 35838; 21pp + Sequence Listing; English.
 XX The invention relates to an isolated nucleic acid detection reagent.
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABL01840-ABL16175) and the encoded proteins
 CC (ABB57737-ABB72072).
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX

SQ Sequence 669 AA;

Query Match 69.0%; Score 29; DB 22; Length 669;
 Best Local Similarity 66.7%; Pred. No. 5.3e+02;
 Matches 6; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 QRIKASKSF 9

:||| |||

Db 61 KRIRLEKSF 69

Search completed: November 6, 2002, 12:05:20
 Job time : 27.7778 secs